

IDC DOCUMENTATION

IDC Processing of Seismic, Hydroacoustic, and Infrasonic Data



Notice

Every effort was made to ensure that the information in this document was accurate at the time of printing. However, the information is subject to change.

Contributors

Greg W. Beall, Science Applications International Corporation
David J. Brown, Science Applications International Corporation
Jerry A. Carter, Science Applications International Corporation
Mark Fisk, Mission Research Corporation
Jeff Hanson, Science Applications International Corporation
Hans Israelsson, Science Applications International Corporation
Richard D. Jenkins, Science Applications International Corporation
David Jepsen, Australian Geological Survey Organisation
Charles N. Katz, Science Applications International Corporation
Ronan J. LeBras, Science Applications International Corporation
Walter Nagy, Science Applications International Corporation
Darrin D. Wahl, Science Applications International Corporation
Jin Wang, Science Applications International Corporation
Xiaoping Yang, Science Applications International Corporation

Trademarks

UNIX is a registered trademark of UNIX System Labs, Inc.

Ordering Information

This document was issued by the Monitoring Systems Operation of Science Applications International Corporation (SAIC) as part of the International Data Centre (IDC) Documentation. The ordering number for this document is SAIC-99/3023, published March 1999. Copies of this document may be ordered by FAX: (619) 458-4993.

This document is cited within other IDC documents as [IDC5.2.1].

IDC Processing of Seismic, Hydroacoustic, and Infrasonic Data

CONTENTS

<u>About this Document</u>	i
■ <u>PURPOSE</u>	ii
■ <u>SCOPE</u>	ii
■ <u>AUDIENCE</u>	ii
■ <u>RELATED INFORMATION</u>	ii
■ <u>USING THIS DOCUMENT</u>	iv
<u>Conventions</u>	vi
<u>Overview</u>	1
<u>Seismic Station Processing</u>	7
■ <u>OVERVIEW</u>	8
■ <u>CHECKING DATA QUALITY</u>	10
■ <u>DETECTING SIGNALS</u>	11
<u>Improving Array Station Snr</u>	11
<u>Improving 3-C Station Snr</u>	18
<u>Detecting</u>	19
■ <u>EXTRACTING SIGNAL CHARACTERISTICS</u>	22
<u>Estimating Arrival Time</u>	22
<u>Measuring Amplitude and Period</u>	27
<u>Estimating Azimuth and Slowness</u>	30
■ <u>ASSOCIATING STATION PHASES</u>	33
<u>Determining Signal Types</u>	33
<u>Grouping Signals and Identifying Phases</u>	37
■ <u>LOCATING AND ESTIMATING MAGNITUDES OF SINGLE-STATION EVENTS</u>	41
<u>Hydroacoustic Station Processing</u>	43

■ OVERVIEW	44
■ CHECKING DATA QUALITY	46
■ DETECTING SIGNALS	46
Filtering	47
Detecting	48
■ EXTRACTING SIGNAL CHARACTERISTICS	49
Estimating Time Features	50
Estimating Energy Features	52
Estimating Moment Features	54
Estimating Cepstral Features	54
Estimating Miscellaneous Features	57
■ ASSOCIATING STATION PHASES	57
Determining Signal Type	57
Grouping Signals and Identifying Phases	60
Infrasonic Station Processing	61
■ OVERVIEW	62
■ CHECKING DATA QUALITY	63
■ DETECTING SIGNALS	64
Identifying Periods of Spatial Coherence	64
Detecting	66
Optimizing Detection Processing of Infrasonic Data	67
■ EXTRACTING SIGNAL CHARACTERISTICS	68
Estimating Arrival Time	68
Estimating Azimuth and Slowness	68
■ GENERATING MAXIMUM COHERENCE TRACE	69
■ ASSOCIATING STATION PHASES	69
Determining Signal Type	69
Grouping Signals and Identifying Phases	71
Network Processing	73
■ OVERVIEW	74
■ DEVELOPING GRIDS	78

Surface Cells	78
Depth Cells	80
Grid Contents	80
■ ASSOCIATING PHASES BASED ON GRID POINTS	82
Identifying Driver Arrivals	82
Searching for Corroborating Arrivals	84
Applying Event Definition Criteria	86
■ EXTRACTING LARGE EVENTS	89
■ ANALYZING SPLITS AND REDUNDANCY	90
Eliminating Split Degenerate Hypotheses	90
Generating Alternate Regional Group Hypotheses	91
Eliminating Redundant Event Hypotheses	91
■ ANALYZING LOCATION AND OUTLIERS	93
Analyzing Outliers	93
Redundancy Testing	94
Applying Event Confirmation Criteria	94
Hydroacoustic Blockage	95
■ RESOLVING CONFLICTS	96
Applying Cluster Analysis	96
Applying Association-based Conflict Resolution	97
■ PREDICTING PHASES	106
Predicting Nondefining Phases	106
Predicting Defining Phases	106
Hydroacoustic Phase Blockage	107
■ CHECKING EVENT CONSISTENCY	107
■ TAGGING ARRIVALS FOR SUBSEQUENT PROCESSING	109
Late Arriving Data	109
Auxiliary and Analyst Reviewed Arrivals	110
Hydroacoustic Arrivals	110
■ TIME WINDOWING AND TIME STEPPING	111
■ RETRIEVING AUXILIARY SEISMIC STATION DATA	113
Ranking Stations	114

User-specified Ranking	114
Requesting Intervals of Data	115
Optimizing Auxiliary Seismic Data Request Parameters	116
Event Location	119
■ OVERVIEW	120
Applications	122
■ INPUTTING DATA	123
Inclusion of Azimuth and Slowness	124
■ ESTABLISHING INITIAL (SEED) LOCATIONS	125
■ PREDICTING TRAVEL-TIMES, AZIMUTHS, AND SLOWNESSES	127
Predicting Seismic Travel-time, Azimuth, and Slowness	127
Predicting Hydroacoustic Measurements	135
Predicting Infrasonic Measurements	137
■ USING LEAST-SQUARES TO INVERT DATA	138
■ EVALUATING SOLUTIONS	140
Stability	140
Divergence Tests	141
Convergence Tests	141
Maximum Iteration Test	142
■ UPDATING EVENT LOCATIONS	143
■ ESTIMATING ERRORS	143
Analyst Review	145
■ OVERVIEW	146
■ REVIEWING AUTOMATED PROCESSING RESULTS	148
Preparing for Analysis	148
Selecting Events for Review	148
Analyzing Events	148
Checking Event Quality	151
Scanning for Aftershocks	151
Checking Analysis Quality	152
Scanning for Missed Events	152

Checking Added Event Quality	152
Checking Quality of Bulletin	152
■ DEFINING EVENTS	153
■ TECHNIQUES	155
Enhancing Signals	155
Identifying Phases	156
Noise and Data Irregularities	156
Categorizing and Evaluating Events	157
Post-analysis Processing	159
■ OVERVIEW	160
■ ESTIMATING EVENT CHARACTERISTICS	162
Estimating Magnitudes	162
Estimating Teleseismic P-wave Complexity	165
Estimating Regional Phase Time-domain Amplitudes	165
Estimating Energy Ratio	167
Estimating Spectral and Cepstral Characteristics	167
Estimating Third Moment of Frequency	168
Estimating Time-frequency Characteristics	169
Estimating First Motion	169
■ APPLYING STANDARD EVENT SCREENING CRITERIA	170
Screening Events by Depth	170
Screening Events by $M_s:m_b$	171
Combining Screening Scores	172
Screening Category Definitions	172
Categorizing Events by Location	173
Algorithms	175
■ OVERVIEW	177
■ FREQUENCY FILTERS	177
Infinite Impulse Response Filters	177
Finite Impulse Response Filters	185
■ INTERPOLATION	186

Closest Sample Interpolation	186
Linear Interpolation	187
Cubic Interpolation	187
Two-dimensional Interpolation	188
■ BEAM GENERATION	191
■ NORMALIZED CROSS-CORRELATION FUNCTIONS	193
■ F-STATISTICS	196
■ STA/LTA DETECTORS	198
Seismic and Hydroacoustic STA/LTA Detectors	198
Infrasonic STA/LTA Detectors	201
■ SEISMIC ARRIVAL-TIME ESTIMATION VIA AKAIKE INFORMATION CRITERION	203
■ FREQUENCY-WAVENUMBER ANALYSIS	206
Error Estimation	209
■ MODIFIED CRAMER-RAO VARIANCE ESTIMATION	212
■ POLARIZATION ANALYSIS	215
■ CEPSTRAL ANALYSIS	218
■ NOISE SPECTRUM EQUALIZATION	225
■ PROBABILITY WEIGHTED TIME	227
Peak Probability	227
Arrival-time Estimation	228
■ NEURAL NETWORKS	229
■ BAYESIAN INFERENCE	232
■ MAGNITUDES	233
Estimating Body Wave Magnitudes	234
Estimating Local Magnitudes	237
Estimating Surface Wave Magnitudes	239
■ DETECTION PROBABILITY	240
■ EVENT SCREENING STATISTICS	243
Depth	243
Surface Wave Magnitude:Body Wave Magnitude	245

<u>Combined Screening Score</u>	246
■ <u>AUTOMATIC ASSOCIATION CHI-SQUARED TEST</u>	247
<u>Mathematical Formalism</u>	247
<u>Chi-square Test</u>	251
■ <u>EVENT LOCATION ERROR ESTIMATION</u>	252
<u>References</u>	257
<u>Glossary</u>	G1
<u>Index</u>	I1

IDC Processing of Seismic, Hydroacoustic, and Infrasonic Data

FIGURES

<u>FIGURE 1.</u>	<u>SYSTEM FUNCTIONAL MODEL</u>	3
<u>FIGURE 2.</u>	<u>S/H/I DATA PROCESSING FLOW</u>	5
<u>FIGURE 3.</u>	<u>STATION PROCESSING OF SEISMIC DATA</u>	9
<u>FIGURE 4.</u>	<u>PLANE WAVEFRONT TRAVERSING ARRAY STATION</u>	12
<u>FIGURE 5.</u>	<u>RESPONSE OF GENERIC NINE-ELEMENT SEISMIC ARRAY</u>	14
<u>FIGURE 6.</u>	<u>BEAMFORMING DELAY AND SUM</u>	15
<u>FIGURE 7.</u>	<u>ONSET TIME ESTIMATION WINDOWS</u>	25
<u>FIGURE 8.</u>	<u>AMPLITUDE AND PERIOD ESTIMATION FOR MB</u>	28
<u>FIGURE 9.</u>	<u>F-K ANALYSIS OF TYPICAL P ARRIVAL AT ASAR</u>	31
<u>FIGURE 10.</u>	<u>STATION PROCESSING OF HYDROACOUSTIC DATA</u>	45
<u>FIGURE 11.</u>	<u>EXAMPLES OF T AND H PHASES</u>	47
<u>FIGURE 12.</u>	<u>TERMINATION TIME ESTIMATION</u>	51
<u>FIGURE 13.</u>	<u>DETRENDED SPECTRUM AND CEPSTRUM FOR H PHASE</u>	55
<u>FIGURE 14.</u>	<u>DETRENDED SPECTRUM AND CEPSTRUM FOR T PHASE</u>	56
<u>FIGURE 15.</u>	<u>STATION PROCESSING OF INFRASONIC DATA</u>	63
<u>FIGURE 16.</u>	<u>STRATOSPHERIC AND THERMOSPHERIC ARRIVALS</u>	70
<u>FIGURE 17.</u>	<u>NETWORK PROCESSING MODEL</u>	75
<u>FIGURE 18.</u>	<u>THREE-DEGREE GRID COVERAGE OVER NORTHERN EUROPE</u>	79
<u>FIGURE 19.</u>	<u>MAP OF WESTERN PACIFIC DEPTH CELLS FOR GA GRID</u>	81
<u>FIGURE 20.</u>	<u>THEORETICAL AND OBSERVED SLOWNESS VECTORS FOR GA DRIVER</u>	83
<u>FIGURE 21.</u>	<u>TWO-STATION LOCATION IN GA ASSOCIATION</u>	85
<u>FIGURE 22.</u>	<u>INTERSECTION OF BEAM WITH SURFACE OF EARTH</u>	87
<u>FIGURE 23.</u>	<u>LINKS BETWEEN GA PRELIMINARY EVENT HYPOTHESES AND ARRIVALS</u>	92
<u>FIGURE 24.</u>	<u>EXAMPLE OF MEMBERSHIP FUNCTION (M_k)</u>	99
<u>FIGURE 25.</u>	<u>GA EVENT HYPOTHESES WITH ARRIVALS IN CONFLICT</u>	103
<u>FIGURE 26.</u>	<u>ASSOCIATION-BASED CONFLICT RESOLUTION PROCEDURE IN GA</u>	105

<u>FIGURE 27.</u>	<u>TIME-WINDOW PARAMETERS FOR GA DATA PROCESSING</u>	111
<u>FIGURE 28.</u>	<u>EVENT LOCATION PROCESSING</u>	121
<u>FIGURE 29.</u>	<u>SSSC MODEL PARAMETERIZATION</u>	131
<u>FIGURE 30.</u>	<u>SLOWNESS/AZIMUTH STATION CORRECTION FOR TXAR</u>	133
<u>FIGURE 31.</u>	<u>BLOCKAGE FILE FOR STATION WK30</u>	137
<u>FIGURE 32.</u>	<u>S/H/I INTERACTIVE ANALYSIS PROCESS</u>	147
<u>FIGURE 33.</u>	<u>POST-ANALYSIS PROCESSING</u>	161
<u>FIGURE 34.</u>	<u>NEURAL NETWORK</u>	230
<u>FIGURE 35.</u>	<u>INDIVIDUAL NODE IN NEURAL NETWORK</u>	231
<u>FIGURE 36.</u>	<u>GEOMETRY FOR AUTOMATIC ASSOCIATION CHI-SQUARE TEST</u>	251

IDC Processing of Seismic, Hydroacoustic, and Infrasonic Data

TABLES

<u>TABLE I:</u>	<u>DATA FLOW SYMBOLS</u>	vi
<u>TABLE II:</u>	<u>TYPGRAPHICAL CONVENTIONS</u>	vii
<u>TABLE III:</u>	<u>TERMINOLOGY</u>	viii
<u>TABLE 1:</u>	<u>SOFTWARE USED IN STATION PROCESSING</u>	10
<u>TABLE 2:</u>	<u>SAMPLE ARRAY STATION DETECTION BEAM SET</u>	17
<u>TABLE 3:</u>	<u>SAMPLE 3-C STATION CHANNEL SET</u>	18
<u>TABLE 4:</u>	<u>HORIZONTAL VELOCITIES OF SEISMIC SIGNAL TYPES</u>	33
<u>TABLE 5:</u>	<u>DEFAULT RULES FOR ASSIGNING SIGNAL TYPES</u>	34
<u>TABLE 6:</u>	<u>SOFTWARE USED IN STATION PROCESSING</u>	45
<u>TABLE 7:</u>	<u>TYPICAL FILTERS USED IN HYDROACOUSTIC PROCESSING</u>	48
<u>TABLE 8:</u>	<u>FEATURES EXTRACTED FROM HYDROACOUSTIC DATA</u>	49
<u>TABLE 9:</u>	<u>FEATURES USED IN DEFAULT RULES FOR HYDROACOUSTIC PHASE IDENTIFICATION</u>	58
<u>TABLE 10:</u>	<u>SOFTWARE USED IN INFRASONIC STATION PROCESSING</u>	63
<u>TABLE 11:</u>	<u>TYPICAL FILTERS USED IN INFRASONIC SIGNAL PROCESSING</u>	65
<u>TABLE 12:</u>	<u>SOFTWARE USED IN NETWORK PROCESSING</u>	75
<u>TABLE 13:</u>	<u>PHASE WEIGHTS FOR EVENT DEFINITION IN SEL1, SEL2, AND SEL3</u>	86
<u>TABLE 14:</u>	<u>SOFTWARE USED IN EVENT LOCATION PROCESSING</u>	121
<u>TABLE 15:</u>	<u>PHASE WEIGHTS FOR EVENT DEFINITION IN REB, SEB, AND SSEB</u>	154
<u>TABLE 16:</u>	<u>SOFTWARE USED IN POST-ANALYSIS PROCESSING</u>	161
<u>TABLE 17:</u>	<u>SEISMIC EVENT-SCREENING CATEGORIES AND CRITERIA</u>	173
<u>TABLE 18:</u>	<u>DEFINITIONS OF LOCATION CATEGORIES</u>	173

About this Document

This section describes the organization and content of the document and includes the following topics:

- [Purpose](#)
- [Scope](#)
- [Audience](#)
- [Related Information](#)
- [Using this Document](#)

About this Document

PURPOSE

This document describes how the International Data Centre ([IDC](#)) processes seismic, hydroacoustic, and infrasonic data.

SCOPE

IDC processing of seismic, hydroacoustic, and infrasonic (S/H/I) data begins after the accumulation of data at the IDC and includes the algorithms, computations, and criteria used to produce the S/H/I data products available from the IDC. The document does not describe the detailed procedures for collecting the data, interactive analysis, data migration, or distributing the data products. These topics are described in sources cited in [Related Information](#), below.

AUDIENCE

This document is intended for scientists, technicians, and managers involved in the operation, maintenance, or use of the IDC and the data products that it provides.

RELATED INFORMATION

IDC documents that might be considered companion documents in the discussion of the seismic, hydroacoustic, and infrasonic processing techniques include the *Database Schema* [\[IDC5.1.1Rev1\]](#) and *Analyst Instructions for Seismic, Hydroacoustic, and Infrasonic Data* [\[IDC6.2.5\]](#). The software that does the actual processing is described in the documentation supporting Automatic Processing software of the

IDC (see category 7.1 of the Roadmap). Manual pages for these programs are documented in [\[IDC6.4Rev1\]](#), and users manuals can be found in category 6.5 "Software User Manuals."

The design and engineering of the IDC are described in the System Design and Engineering documentation (see category 4 of the Roadmap).

The system used by the IDC draws from many years of international cooperative development in seismic signal detection and location since the 1958 Geneva expert conference. Specific procedures employed by the IDC evolved more recently from seismic monitoring systems that used data from arrays to detect and locate events automatically. Automated processing systems using regional array data were developed in concert with regional arrays in the early 1980's [\[Myk90\]](#). For example, the RONAPP (Regional ON-line Array Processing Package) program automatically processed data from a single regional array [\[Myk84\]](#) and used azimuth and differences in phase arrival time to locate events using the data from a single array station [\[Bra88\]](#). The network of stations grew as did the system for processing their data. Data from several regional array stations in Fennoscandia and Germany were automatically processed with the Intelligent Monitoring System [\[Bac90\]](#). The automated systems were tested in a seismic monitoring system on a global scale for the first time during the Group of Scientific Experts Technical Test 2 (GSETT-2) between April 22 and June 9, 1991 (42 days) [\[GSE91\]](#). GSETT-3 [\[GSE94\]](#) marked the beginning of continuous operations for the seismic monitoring system. The operations manual [\[GSE95b\]](#) describes the processing used at the IDC for GSETT-3, which started on January 1, 1995 and still continues. Since that document was written, procedures have changed substantially, and hydroacoustic, infrasonic and radionuclide processing capabilities have been added to the system.

Many of the processing techniques and algorithms used at the IDC are described in journal articles that are referenced throughout this document. See ["References" on page 257](#) for a listing of all the sources of information consulted in preparing this document. Articles on seismic monitoring can also be found in the December, 1990, Special Symposia Issue of the *Bulletin of the Seismological Society of America* [\[Rin90\]](#).

USING THIS DOCUMENT

This document is part of the overall documentation architecture for the International Data Center as charted on the Roadmap located on the pages preceding the Table of Contents.

This document is organized as follows:

- [Overview](#)
This chapter is a general overview of the processing sequence used at the IDC.
- [Seismic Station Processing](#)
This chapter describes the station processing techniques used for IMS seismic data.
- [Hydroacoustic Station Processing](#)
This chapter describes the station processing techniques used for IMS hydroacoustic data.
- [Infrasonic Station Processing](#)
This chapter describes the station processing techniques used for IMS infrasonic data.
- [Network Processing](#)
This chapter describes the network processing techniques for associating phases detected at stations with events.
- [Event Location](#)
This chapter describes the location process that leads to new or refined locations for events.
- [Analyst Review](#)
This chapter describes the analysts' steps for reviewing the results obtained through automated processing.

- [Post-analysis Processing](#)

This chapter describes the additional automatic processing that occurs after analyst review. This processing includes the estimation of event characterization parameters and the screening of events that are considered to be consistent with natural phenomena or non-nuclear, man-made phenomena.

- [Algorithms](#)

This chapter describes the mathematical algorithms for many of the processing techniques described in the previous chapters.

- [References](#)

This section lists the sources cited in this document.

- [Glossary](#)

This section defines the terms, abbreviations, and acronyms used in this document.

- [Index](#)

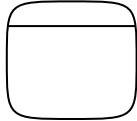




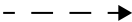
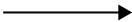

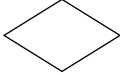

This section lists topics and features provided in this document along with page numbers for reference.

About this Document ▼

Conventions

This document uses graphical, typographical, and naming conventions, as described in Tables I, II, and III. Table I shows data flow symbols used in figures of IDC processes.

TABLE I: DATA FLOW SYMBOLS

Description	Symbol
process	
external source or sink of data (left)	
duplicated external source or sink of data (right)	
data store (left)	
duplicated data store (right)	
control flow	
data flow	
node	
decision	
product	

[Table II](#) shows the typographical conventions followed throughout this document.

TABLE II: TYPOGRAPHICAL CONVENTIONS

Element	Font	Example
database table	bold	dataready
database table and attribute when written in the dot notation		prodtrack.status
headings, figure titles, and table titles		About this Document
matrices and vectors		R_x
attributes of database tables when written separately	<i>italics</i>	<i>status</i>
processes and software units		<i>ParseSubs</i>
user-defined arguments		<i>delete-remarks object</i>
titles of documents		<i>Database Schema</i>
libraries		<i>libloc</i>
computer code and output	courier	>(list 'a 'b 'c)
filenames, directories, and websites		amp.par
text that should be typed in exactly as shown		edit-filter-dialog

**About this ▼
Document**

[Table III](#) defines terms that are used in a specific context in this document. See the Glossary at the end of this document for a more general listing of terms, abbreviations, and acronyms.

TABLE III: TERMINOLOGY

Term	Description
characteristic	number included in event lists or bulletins that identifies a quality of a signal or event
feature	measurement made from waveform data
incoherent	not coherent
phase measurement	observation of onset time, azimuth, and slowness for a particular phase
raw data channel	channel as it is transmitted from a station (unprocessed beyond what was done at the station)
rectilinearity	bounded measure that characterizes the degree to which particle motion of a three-component (3-C) seismic signal is along a straight line

Overview

This chapter provides a high-level overview of the techniques by which the IDC processes seismic, hydroacoustic, and infrasonic data. The sequence of processing steps is provided in this overview. Detailed descriptions of the individual processing steps are described in later chapters.

Overview

The objective of S/H/I processing is to produce event lists and bulletins that describe the S/H/I events that occurred in a given period. This objective is achieved by a combination of automatic processing and interactive analysis by human specialists. Automatic processing is used to produce the Standard Event Lists ([SEL1](#), [SEL2](#), and [SEL3](#)), which are progressively more delayed from real-time, and thus contain progressively more refined event characterization as more data arrive at the [IDC](#). Analysts examining the results contained in the best automated event list ([SEL3](#)) review and modify the results as necessary. The quality of the reviewed events is checked, and the events are placed in the Reviewed Event Bulletin ([REB](#)). Subsequently, additional automatic processing generates the Standard Event Bulletin (SEB), the Standard Screened Event Bulletin (SSEB), and the Executive Summary.

To monitor a 24-hour period, all of the data that might contain signals from events within the day must be processed. Because the signals from events travel at different speeds from the source location through the earth, oceans, and atmosphere to the stations, the amount of data needed to process a day of events is dependent on the path of the signals and may extend several hours into the following day.

Seismic stations record signals that have travelled through the solid earth. The velocities are relatively high, and the IDC can obtain enough data to process a terrestrial event located anywhere in the solid earth in less than an hour. Hydroacoustic stations record signals that have travelled more slowly through the oceans. Approximately two hours are required for the IDC to obtain enough data to locate an event in the ocean using hydroacoustic data alone. Infrasonic stations record signals that have travelled through the atmosphere. Propagation velocities are much slower in the atmosphere, and several hours are required for the IDC to obtain enough data to process an atmospheric event.

The workflow of IDC processing can be split into four phases as shown in [Figure 1](#); station processing, network processing, interactive analysis, and post-analysis processing. Station processing is performed automatically and independently on each of the data types obtained from the International Monitoring System (IMS). The main task of station processing is to detect signals in the data and determine from extracted signal parameters if the signal is from an event or is noise. Network processing uses the results of station processing from many different stations to locate events. Network processing is performed at several times in the processing sequence and includes progressively more data each time. Interactive analysis is performed by experts who review the automated results and make changes as appropriate. They also search for events that may have been missed in the automated processing. Post-analysis processing is automatic and computes event characteristics that are then used for event screening.

Each of these major functions uses the operations database to pass information from one process to the next. When all of the processing has been completed, the database information is migrated to an archival database.

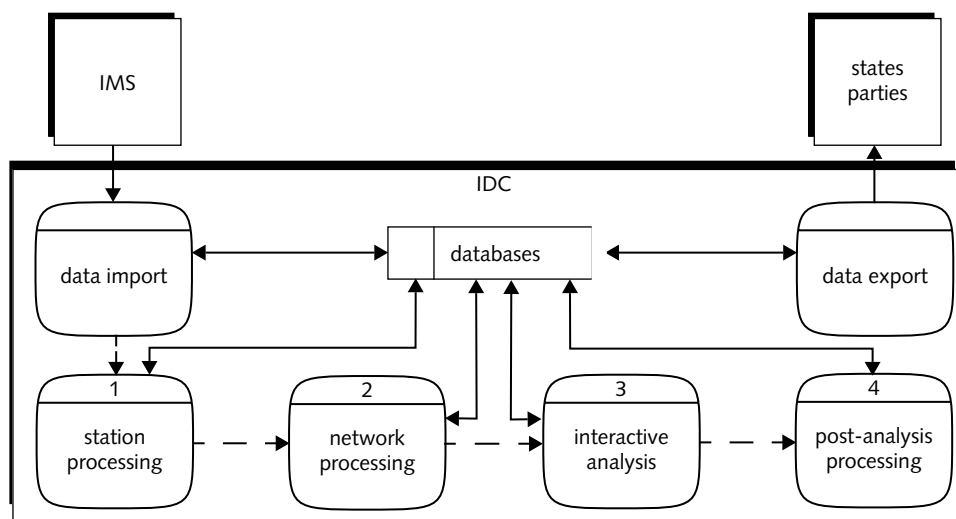


FIGURE 1. SYSTEM FUNCTIONAL MODEL

Overview ▼

The sequence of processing steps for the different types of data is based upon when the data containing signals from the possible sources arrive at the IDC. The processing sequence is shown in [Figure 2](#).

Data from all three types of sensor arrive at the IDC continuously in near-real time, and station processing is performed whenever a large enough sample of data has been collected (see processes 1, 2, 4, 5, 6, 8, 9, and 10 in [Figure 2](#)). For station processing, approximately 10 minutes of data are processed for the seismic and hydroacoustic data in each run. For infrasonic data, 30 minutes of data are processed. The results of station processing are placed in the Standard List of Signal Detections (SLSD), which is a standard product of the IDC.

Network processing is scheduled at several different times after the “event time”. The first network processing is performed on the primary seismic data and any hydroacoustic data with completed station processing (process 3 in [Figure 2](#)). This occurs approximately two hours after event time, and approximately 20 minutes of signal detections are processed in each step. The results of the first network processing are included in the [SEL1](#). The events in the SEL1 are used to generate requests for additional seismic data from the auxiliary seismic stations.

The second network processing run is performed approximately six hours after event time, again using 20 minutes of signal detections in each step. In addition to the signals from the original run, signals detected in the auxiliary seismic data, late arriving primary seismic data, late arriving hydroacoustic data, and infrasonic data are included in the second run (process 1 in [Figure 2](#)). The product of this network processing is the [SEL2](#).

The final network processing run is performed approximately 12 hours after event time (process 11 in [Figure 2](#)). The delay from real time is dependent on the interactive processing schedule established by the IDC. All available signal detections for the data day are included in the run. The product of this network processing is the [SEL3](#).

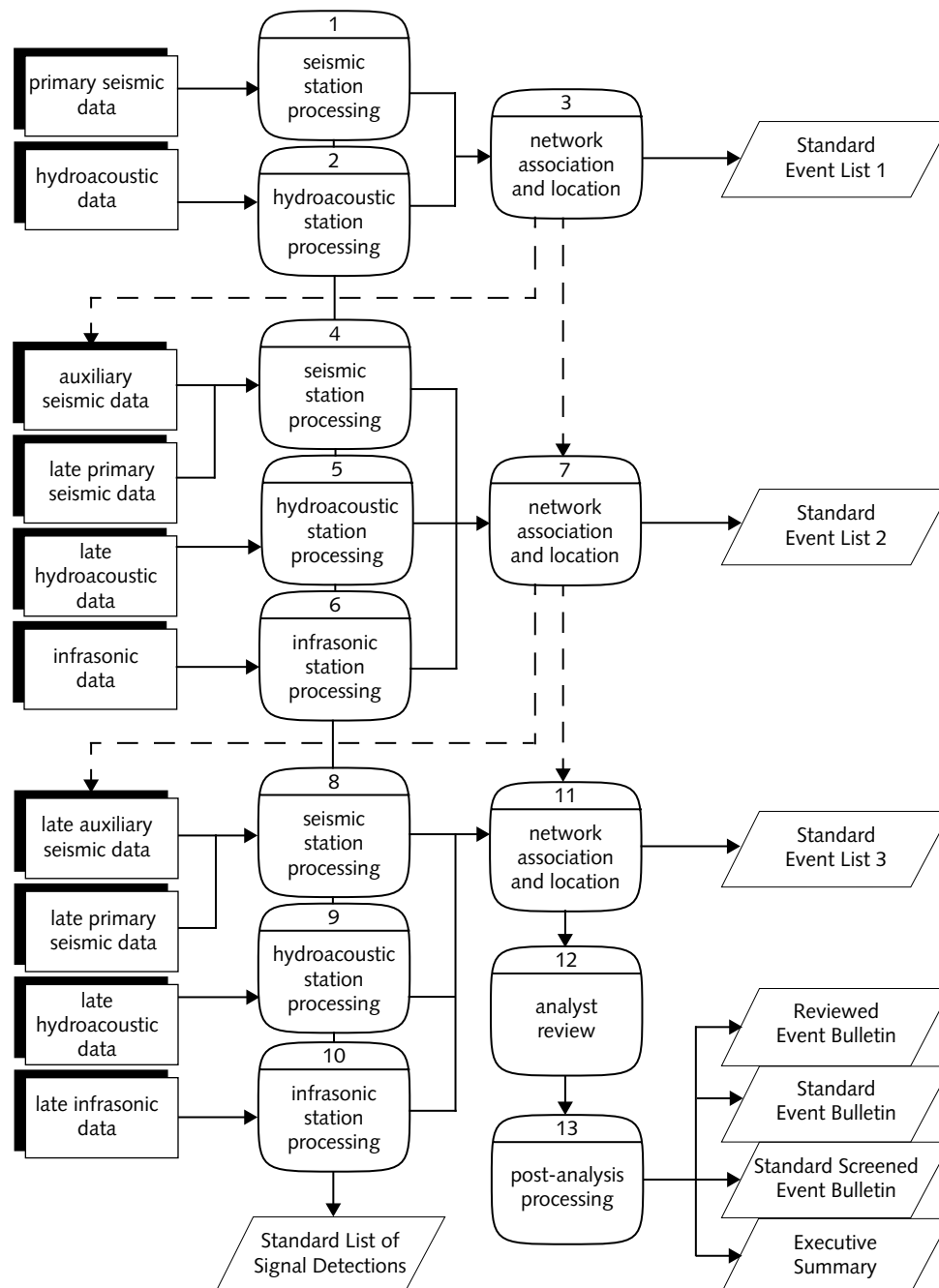


FIGURE 2. S/H/I DATA PROCESSING FLOW

Overview ▼

The results of automated processing are reviewed by expert analysts who delete false events and events that do not meet the event definition criteria, and correct other mistakes made during automatic processing (process 12 in [Figure 2](#)). Analysts also search the data for events that might have been missed by the automatic processing. Analysis is performed on an entire day of data (24 hours beginning at 00:00 Universal Coordinated Time [UTC]). A team of analysts reviews the data; each analyst reviews a few hours of data. Their work is checked for quality by experts in the fields of seismology, hydroacoustics, and infrasonics.

When analysis has been completed, post-analysis processing is applied to the data (process 13 in [Figure 2](#)). This processing includes calculation of amplitudes, periods and signal-to-noise ratio (snr) for added or retimed phases, background noise amplitude and period estimates at stations that did not detect the event, and surface wave magnitudes. These calculations are added to the results of human analysis to produce the REB. The REB is released after human analysis and the subsequent post-analysis processing has been completed (usually 48 hours after the end of the data day).

In addition, post-analysis processing includes the automatic calculation of a series of event characteristics for each event whose location appears in the REB. These characteristics are included in the SEB and are the basis of the screening process. The SEB is the REB with the addition of the calculated event characteristics. Screening criteria are applied to the SEB event characteristics to produce the SSEB. The content and format of the SSEB is the same as the SEB, except that events that are screened out are not included in the SSEB. The SEB and SSEB are released approximately two hours after the REB (about 50 hours after the end of the data day).

Summary information regarding the number of events detected, the numbers of events in the various screening categories, and the status of the IMS network, communications links, and the IDC are included in the Executive Summary. The Executive Summary is released approximately two hours after the REB.

Seismic Station Processing

This chapter describes how the IDC processes data from the seismic IMS stations and includes the following topics:

- [Overview](#)
- [Checking Data Quality](#)
- [Detecting Signals](#)
- [Extracting Signal Characteristics](#)
- [Associating Station Phases](#)
- [Locating and Estimating Magnitudes of Single-station Events](#)

Seismic Station Processing

OVERVIEW

Station processing includes the following tasks:

- checking the quality of the data
- improving the snr of the data
- detecting signals
- estimating the arrival time of the signal
- determining amplitude and period of the signal
- estimating the azimuth and slowness of the signal
- determining the type of signal
- grouping the signals at each station and identifying the phases
- estimating single station locations

These tasks are performed automatically on small time increments of data (10 minutes) as they arrive using the *Detection and Feature Extraction (DFX)* and *Station Processing (StaPro)* applications [\[Wah96a\]](#), [\[Wah96b\]](#). *DFX* checks the data quality, generates channels for detection processing, detects signals in the data, and extracts the signal features from the waveforms. *StaPro* uses these extracted features to determine the signal type, which in turn gets used in network processing (see [“Network Processing” on page 73](#)). The results of station processing are reported in the SLSD.

Two different types of seismic stations are used in the IMS: arrays of single and 3-C seismometers and individual 3-C stations. The waveform data from the multiple station arrays are processed by spectral filtering and beam forming to enhance the snr. The waveform data are subsequently evaluated for the presence of transient energy via a comparison of short-term energy estimates with estimates of

the ambient background energy prior to signal arrival. Ultimately, refined estimates of the arrival azimuth, phase velocity, and emergence angle are computed via frequency-wavenumber spectral analysis. The waveform data from the individual 3-C stations are processed by spectral filtering to enhance the snr. The waveform data from the vertical component are subsequently evaluated for the presence of transient energy via a comparison of short-term energy estimates with estimates of the ambient background energy prior to signal arrival. The waveform data from the two horizontal components are combined incoherently and are also evaluated for the presence of transient energy. Ultimately, refined estimates of the arrival azimuth, phase velocity, and emergence angle are computed via polarization analysis.

The significant station processing steps are shown in [Figure 3](#) and described in the following sections. [Table 1](#) lists the software application that performs each of the station processing steps.

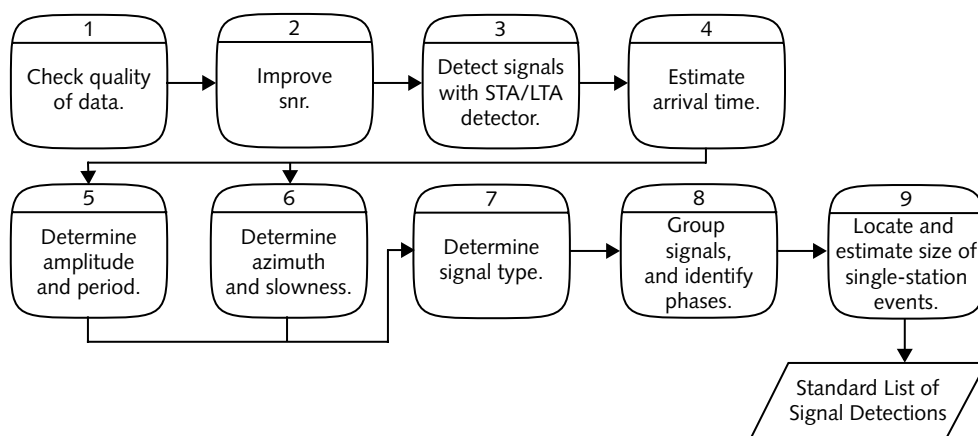


FIGURE 3. STATION PROCESSING OF SEISMIC DATA

TABLE 1: SOFTWARE USED IN STATION PROCESSING

Step	Software	Reference(s)
1 Check quality.	<i>DFX</i>	[Wah96a] , [Wah96b]
2 Improve snr.		
3 Detect signals.		
4 Estimate arrival time.		
5 Estimate amplitude and period.		
6 Estimate azimuth and slowness.		
7 Determine signal type.	<i>StaPro</i>	[IDC7.1.2]
8 Group and identify phases.		
9 Locate single-station event.		

CHECKING DATA QUALITY

All waveform data are stored in the IDC archives as they are received. Prior to signal processing, however, the quality of the data must be checked (process 1 in [Figure 3](#)). Spikes (isolated large amplitude data values), dropouts (missing data), and repeated amplitude values are considered to be problem data. Long-period trends in data can also cause problems in later processing. In some cases, problem data can be repaired through interpolation or “de-trending.” If the data are not repaired, a mask that indicates where the data have quality problems and should not be processed is created.

Problem data must be recognized before they can be repaired or masked. Spikes are recognized by the difference in amplitude relative to surrounding data values. Repeat amplitude values are recognized by the number of consecutive samples with the same amplitude value (usually 4 to 10, depending on the sample rate).

Problem data are repaired if the number of problem samples is small. If the number of consecutive problem samples is less than the number of samples that define a gap (usually four to ten), the problem data are replaced with linearly interpo-

lated data values (see [“Linear Interpolation” on page 187](#)). If the number of consecutive problem samples is larger than the number of samples that define a gap, the data are masked. When a large percentage (usually more than 33%) of the data in the time interval are masked, then the entire interval is considered suspect and will not be processed.

Removing the trend in unmasked data is optional. Trends are removed by fitting a linear trend to the data and then subtracting the linear trend from the data values.

Details on the explicit implementation of quality control can be found in [\[Wah96a\]](#) and [\[Wah96b\]](#).

DETECTING SIGNALS

Low amplitude signals are sometimes difficult to detect in raw (unprocessed) waveform data. The many channels of raw data received from the stations are processed in an effort to improve the snr for the types of signals that are of interest (process 2 in [Figure 3](#)) prior to detection processing (process 3 in [Figure 3](#)).

Improving Array Station Snr

Array stations have spatially distributed sensors at which data are recorded. As a signal traverses the array, it is sensed by the individual elements of the array at different times (see [Figure 4](#)). The effective time delay between any given pair of sensor elements is governed by the distance between sensor elements along the direction of propagation as well as the slowness of the signal. The latter is the inverse velocity in the plane of the sensor array. The array configuration determines how well signals from different azimuths and slownesses can be resolved.

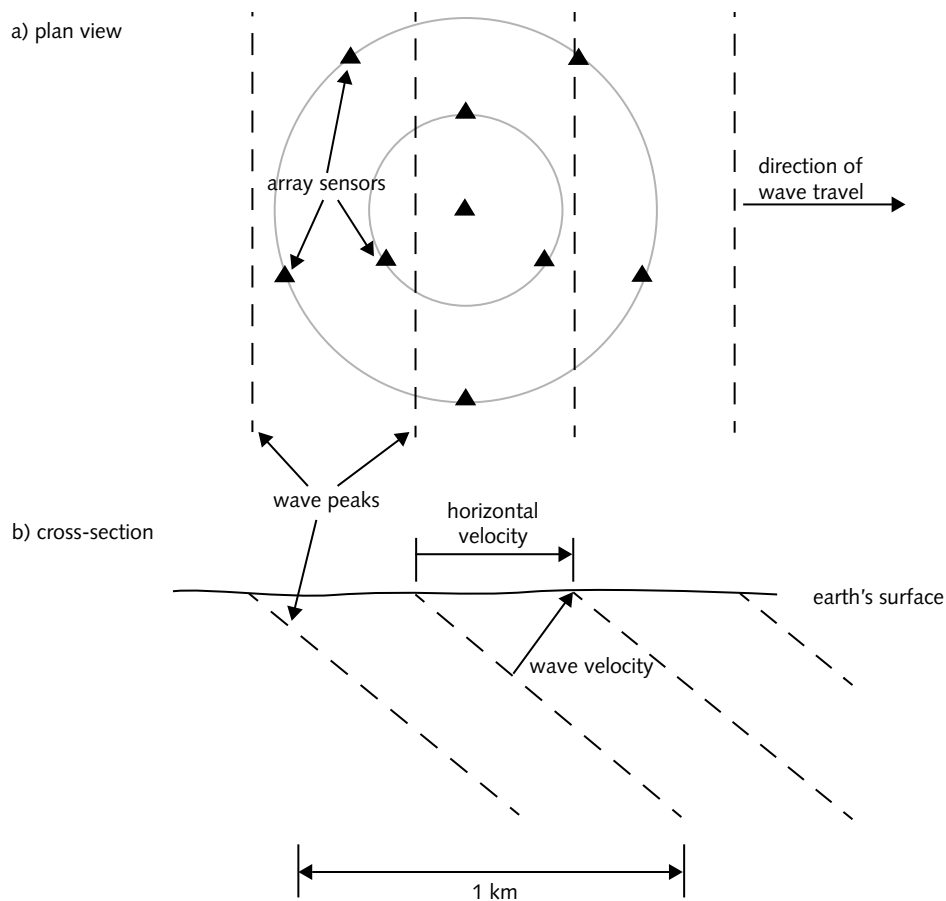


FIGURE 4. PLANE WAVEFRONT TRAVERSING ARRAY STATION

[Figure 5](#) shows the response of a generic nine-element seismic array in coordinates of wavenumber (frequency \times slowness; using wavenumber eliminates the frequency-dependence of this diagram). The high point at the center of the figure indicates that the array is most sensitive to signals travelling vertically. Moving away from the center of the diagram, the response is significantly diminished. The shape of the central peak defines the resolution of the array. As the dimensions of the array and the number of sensors in the array increase, the sharpness and height of the peak also increase.

Array data would not be useful if the only signals that could be detected were travelling vertically. The peak in [Figure 5](#) is shifted to other areas of the wavenumber plane by time shifting the data from the sensors relative to one another prior to summing. In other words, the data from the sensors are combined in various ways to form new data channels that enhance signals travelling in certain directions at certain velocities. A new data channel created through processing is known as a “beam” and the collection of beams used for a particular station is called the “beam set.” Beams may be either coherent or incoherent (rectified), may be “steered” to specific azimuths and horizontal velocities, may be frequency filtered, and may include all or a subset of the raw data channels available from a particular station.

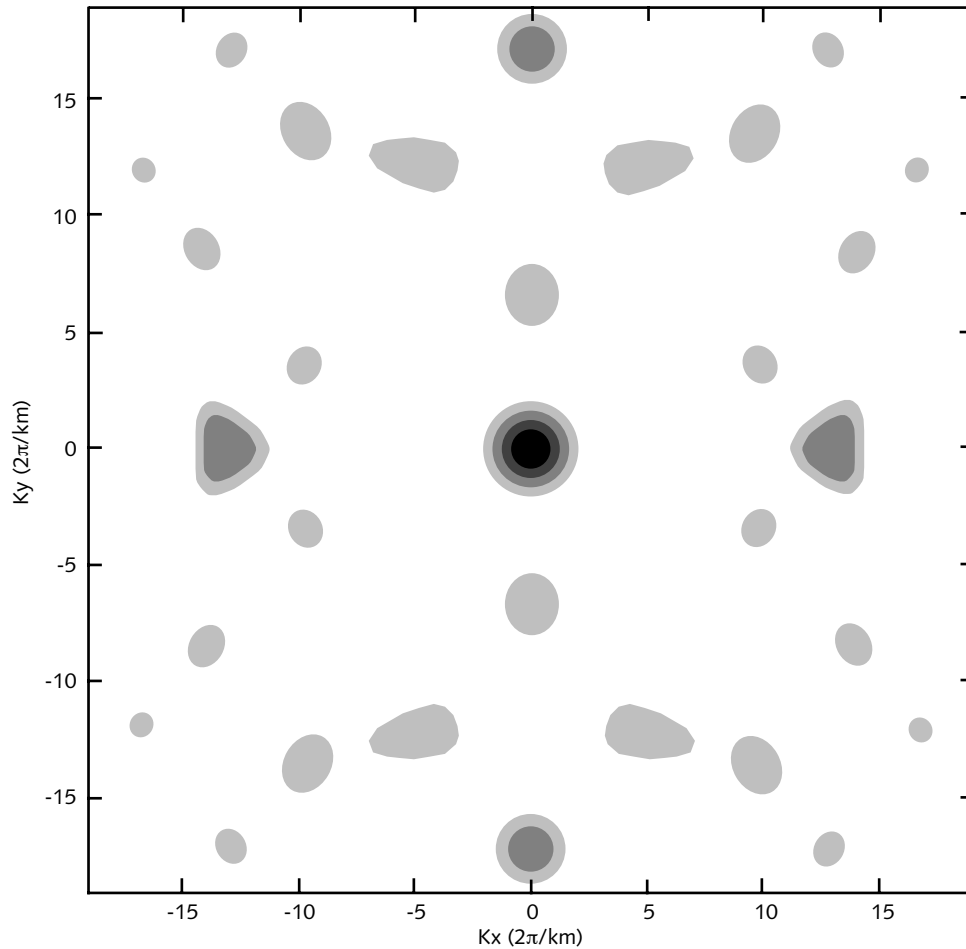
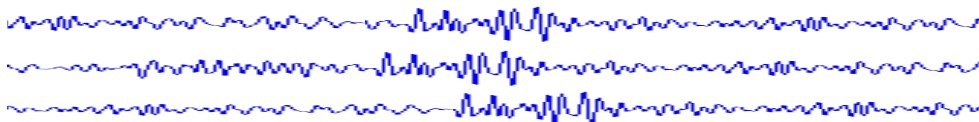


FIGURE 5. RESPONSE OF GENERIC NINE-ELEMENT SEISMIC ARRAY

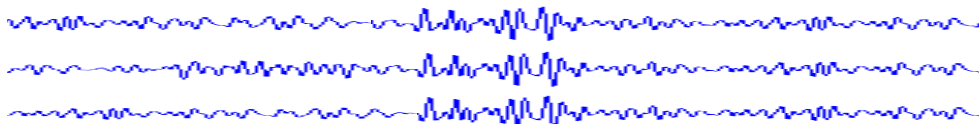
Because signals may travel through the array in any direction, the individual beams of the beam set are “steered” to specific azimuths and slownesses so that the ability to detect signals is fairly uniform in all directions. To steer the beams, the raw data channels of an array are shifted in time relative to one another and summed (see [Figure 6](#)). This process enhances signals travelling along the azimuth and slowness of the steered beam relative to the noise, which is assumed to be

incoherent. The frequency characteristics of seismic signals dictate the frequency bands that will be dominant for specific slownesses. The frequency filters are causal band-pass filters (see [“Infinite Impulse Response Filters” on page 177](#)).

raw data channels



data channels delayed by assumed azimuth and slowness



data channels summed to form beam



FIGURE 6. BEAMFORMING DELAY AND SUM

If the noise is not completely incoherent, then additional improvements in the snr can be obtained in array beams by selecting subsets of the raw data channels to use in the beam. The raw data channels used to form such beams are chosen so that the coherent noise is cancelled, thereby improving the snr [\[Myk90\]](#).

“Incoherent” beams are useful for detecting horizontally polarized signals and for measuring some amplitudes. These beams are formed by frequency filtering the data, shifting the filtered channels, and then summing the absolute value of the filtered channel values.

The beam set used for a particular station is based on experience with the signals and noise conditions at that station and the station configuration (array geometry, sensor orientation). The beam set may change as experience is gained with the data from the station. References describing the theory of detecting signals from seismic array data include [\[Myk84\]](#), [\[Kvæ89\]](#), and [\[Myk90\]](#).

Optimizing Beam Parameters

Optimizing parameters for improving the snr ensures that signals of interest will not be missed during station processing.

For array stations, the beam set must cover the range of slownesses in which seismic signals are detected. Each beam has a “response” in azimuth and slowness that is based on the dimension of the array, the number of sensors in the array, the distribution of the sensors within the array, and the frequency band of the data used in the beam. A beam response is maximum at the azimuth and slowness used to steer the beam and tapers away as azimuth and slowness change. The beam responses are used to establish a beam set that covers the range of the frequencies, azimuths, and slownesses that contain seismic signals of interest.

One sequence for establishing an array beam set is as follows:

1. Compute the “array pattern” as the two-dimensional spatial Fourier transform of the weighted (usually unit weight) sensor locations.
2. Determine the wavenumber resolution and alias frequency in the wavenumber domain from the array pattern.
3. Choose a set of overlapping frequency bands. For the regional arrays the frequency bands include: 0.75–2.25, 1.0–3.0, 2.0–4.0, and 3.0–6.0 Hz.
4. Set fixed slowness values for the regional phases (Pn: 0.1237 s/km; Pg: 0.1640 s/km; Sn: 0.2225 s/km; Lg: 0.2857 s/km; Rg: 0.3334 s/km).
5. Determine the slowness interval of each frequency band based on the wavenumber resolution for teleseismic phases.
6. Derive the steered beam pattern from the azimuth resolution, which depends on frequency, slowness and the wavenumber resolution.

[Table 2](#) provides an example of a beam set used for detection for a generic nine-element array of vertical sensors with an array aperture of 2 km. The beam set contains 99 beams.

The beams listed in [Table 2](#) are used only for detection. Other beams for amplitude estimation, analysis, and threshold monitoring may also be included in the beam set.

TABLE 2: SAMPLE ARRAY STATION DETECTION BEAM SET

Purpose	Type	Frequency Filter	Azimuths	Slowness (s/km)
teleseismic phase detection	coherent	0.75–2.25 Hz	0	0.0000
			0, 60, 120, ...	0.1061
	coherent	1.00–3.00 Hz	0	0.0000
			0, 60, 120, ...	0.0796
	coherent	2.00–4.00 Hz	0	0.0000
			0, 60, 120, ...	0.0531
			30, 90, 150, ...	0.1061
	coherent	3.00–6.00 Hz	0	0.0000
			0, 60, 120, ...	0.0354
			30, 90, 150, ...	0.0707
			0, 40, 80, ...	0.1061
	coherent	4.00–8.00 Hz	0	0.0000
Pn phase detection	coherent	1.00–3.00 Hz	30, 90, 150, ...	0.1237
Pg phase detection	coherent	2.00–4.00 Hz	0, 40, 80, ...	0.1640
Sn phase detection	coherent	1.00–3.00 Hz	0, 40, 80, ...	0.2225
Lg phase detection	coherent	2.00–4.00 Hz	0, 24, 48, ...	0.2857
Rg phase detection	coherent	0.75–2.25 Hz	0, 36, 72, ...	0.3334

Improving 3-C Station Snr

Three-component seismic stations record three orthogonal directions of ground motion from a single site. The vertical component of motion is used primarily for detecting vertically polarized motion (such as P phases), and the horizontal components are used primarily for detecting horizontally polarized motion (such as S phases). Frequency filtering is the primary means of improving the snr of 3-C data. Vertical data are frequency filtered using a causal band-pass filter into a number of channels that cover the range of frequencies expected for vertically polarized phases. The two horizontal channels are frequency filtered in bands that cover the range of frequencies expected for horizontally polarized phases and summed incoherently to form channels for detection.

Optimizing 3-C Snr Parameters

Optimizing parameters for improving the snr ensures that signals of interest will not be missed during station processing.

For 3-C stations, the data channels used for detection are fairly standard because the configuration of 3-C station sensors is fairly standard. Vertical component data are frequency filtered to form channels with overlapping frequency bands, and horizontal component data are frequency filtered and combined incoherently. Two channels are also included for making amplitude measurements. [Table 3](#) shows a sample channel set for a 3-C station.

TABLE 3: SAMPLE 3-C STATION CHANNEL SET

Purpose	Type	Frequency Filter	Channels
P phase detection	coherent	0.50–1.50 Hz	vertical
		1.00–2.00 Hz	vertical
		1.50–3.00 Hz	vertical
		2.00–4.00 Hz	vertical
		3.00–6.00 Hz	vertical
		4.00–8.00 Hz	vertical

TABLE 3: SAMPLE 3-C STATION CHANNEL SET (CONTINUED)

Purpose	Type	Frequency Filter	Channels
S phase detection	incoherent	0.50–1.50 Hz	horizontals
		1.00–2.00 Hz	horizontals
		1.50–3.00 Hz	horizontals
		2.00–4.00 Hz	horizontals
		4.00–8.00 Hz	horizontals

Detecting

Signals in the processed data channels are detected with an algorithm that computes the ratio of the short-term energy to the long-term energy (STA/LTA) (process 3 in [Figure 3](#)). Details of the STA/LTA algorithm are described in [“Algorithms” on page 175](#). A STA/LTA threshold is established for each detection channel that, when exceeded, identifies a potential “detection.” A detection is declared when the STA/LTA thresholds for a specified number of channels (typically one) have been exceeded. When the STA/LTA is above the threshold for a channel, that channel is in a “detecting state.” A channel in a detecting state remains in that state until the STA/LTA value stays below the threshold for a specified time interval.

When STA/LTA thresholds are exceeded for numerous channels at nearly the same time, the channel from which some of the signal characteristics will be estimated should be identified. The channel with the largest STA/LTA value is defined as the “detecting channel.” The STA and STA/LTA values of the detecting channel are entered into the **detection** database table as the *stav* and *snr* columns, respectively (see [IDC5.1.1Rev1](#)). The time of the detection (*time* in the **arrival** and **detection** database tables) is automatically estimated via the Akaike Information Criterion described in [“Extracting Signal Characteristics” on page 22](#).

Optimizing Detection Parameters

The goal of optimizing the detection processing is to achieve a judicious balance between detections that are false alarms (not from real signals) and signals that are missed. The parameter that must be optimized is the STA/LTA threshold for each channel.

The basic procedure is to set the STA/LTA thresholds for each channel to a low level and process “control” data that have been reviewed by analysts. The test detection results are compared to the reviewed event phases and used to estimate the STA/LTA thresholds that would result in a specific percentage of detections being associated by analysts (association rate). The percentage of the detections associated by analysts that were missed (missed rate) using these STA/LTA thresholds is then estimated. The process is repeated until a relationship between the association rate and the missed rate is developed.

The average times required by analysts to add a phase and to review detected phases are considered when deciding the proper balance between the association rate and the missed rate. The analyst time required to scan for an event that was missed in automatic processing due to a missed detection must be considered. The STA/LTA thresholds that come closest to the desired rates is used.

The steps for optimizing the STA/LTA detection thresholds are as follows:

1. Run the *DFX* application using a STA/LTA threshold of 3.0 for all channels on “control” data that have been reviewed by analysts.
2. Determine the relationships between the number and snr of detections for each channel.
3. Determine the relationships between the number and snr of associated phases for each channel.
4. Determine the relationships between the number and snr of added (missed) phases for each channel.
5. Set an objective for the association rate for all channels (around 10%).

6. If the existing STA/LTA threshold for channels with the same slowness and frequency band has an association rate that is lower than the objective, increase the thresholds of the channels to a value for which the objective is achieved.

If the existing STA/LTA threshold for channels with the same slowness and frequency band has an association rate that is higher than the objective, decrease the thresholds of the channels to a value for which the objective is achieved (or the percentage of added phases becomes zero).

7. Repeat steps 4 and 5 until a relationship between the association rate and the added rate is obtained.
8. Choose a point where the added rate is lower than 20% (first priority) and the association rate is about 10% (second priority).
9. Test the modified recipe by running the *DFX* application.

Steps 1–8 in the preceding procedure are based on a single processing run of the *DFX* application. When detections are theoretically eliminated from one channel by increasing the STA/LTA threshold, they may still be detected on other channels. This effect is known as “channel jumping” and no account of this effect is made as it is difficult to predict. Testing the modified recipe in Step 9 is thus crucial. Some adjustment to the STA/LTA thresholds may be needed to reduce the effect of channel jumping.

The detections achieved automatically in pipeline operations are simultaneously written to both the **arrival** and **detection** tables of the database. Each **detection** table entry has many fields in common with the fields of the corresponding **arrival** table entry. The **detection** table entries remain unaltered after they are written to the table. In contrast, the **arrival** table entries, particularly those concerning onset time, *fk* estimated parameters, polarization attributes (where applicable), and amplitude may be subsequently modified by analyst revisions.

EXTRACTING SIGNAL CHARACTERISTICS

The characteristics of seismic signals include arrival time, amplitude, period, azimuth, and slowness (processes 4 through 6 in [Figure 3](#)). These are included in event lists and bulletins. Signal characteristics are either estimated directly from the data or derived from “features” of the data that are measured and stored in the IDC database.

Estimating Arrival Time

Arrival times for the onset of seismic phases are used in both constructing and locating seismic events. Their accuracy directly impacts the quality of many downstream products of the system. Seismic phases are first detected by a rudimentary although highly configurable snr method that is based on short-term and long-term averaging. The algorithm is designed for reliability of detection but provides only a crude estimate of its actual onset time. The initial estimate of detection time is temporarily stored in *DFX* memory. A time window is constructed about this initial estimate that acts as a seed for the more sophisticated onset time estimation based on auto-regressive modeling of the phase (signal) energy or the background noise energy or both. The *time* attribute of both the **arrival** and the **detection** table is set to this value. The *time* value in the **arrival** table may be reset by the analyst in the course of interactive analysis. An uncertainty for the measurement of the detection time is computed from the snr and stored in the *deltim* attribute of the **detection** and **arrival** tables. This value is unchanged in the onset time revision procedure (see [“Detecting Signals” on page 11](#) for details regarding detection times).

Akaike Information Criterion (AIC)

The problem of estimating the onset time of seismic phases has been approached by many authors from an information theoretical perspective. Their methods differ in the details but are all based on evaluating the differences between prior statistical models of the noise and signal processes and the observed statistics of a real

time-series containing energy from either or both. Entropy or entropy-like functions, which can be used to measure the randomness or information in a process or a series of observations, are evaluated for a fixed waveform segment about the seed detection to quantify this comparison. The Akaike Information Criterion (AIC) is a so-called generalized entropy that provides a measure of the information difference between a process and a set of observations.

In its most basic form the AIC is expressed in terms of the pre-determined probability density function (PDF) of the generating process and the empirical PDF of the observations themselves. However, its application is greatly simplified by describing the process statistics by using a predictive model and examining the residual between its predictions and the data. Auto-regressive (AR) prediction models are the customary choice, in keeping with the causal nature of time, although in principle forward-looking models can be used. In theory, the predictor will account for all of the predictable (non-white) energy in a realization of the process. Residuals from its predictions thus have a flat or white pre-determined power spectrum. For this series, the AIC reduces to a linear combination of the residual log-variances plus terms that are effectively constant for a given waveform segment. The particular linear combination depends on which predictor model, noise or signal, is applied to what part or parts of the segment. The mathematical formulation used here is presented in the following section.

The theoretical validity of the AIC is derived in part from assumptions that the noise and signal processes are statistically distinct, stochastic and stationary, and that the transition between the regimes with and without signal is instantaneous. Under such conditions, the expected value of the AIC increases monotonically away from the point of transition between the processes so that a unique minimum in the function determines the time of onset. The AIC itself is a statistical entity, so this, an unambiguous minimum, will not necessarily be realized even if the assumptions are valid for an actual waveform. To the degree to which these assumptions are invalid, the AIC function departs from this ideal shape; local minima appear, the global minimum becomes less sharp and compelling, and the overall method is less effective. Stationarity is not generally true of seismic noise but it is usually a decent approximation over short periods of time. Alternatively, stationarity is usually a poor assumption for seismic signals; signals can be impul-

sive or emergent, have long, short or no coda, and may vary in intensity and frequency. For small events, the event statistics can converge on the noise statistics, but even for larger events, an emergent arrival may result in a degraded AIC shape and more ambiguous onset determination.

AIC Method

Many variations exist on how the AIC is constructed and applied. A number of these variations are described in [\[Kam92\]](#) and subsequently used by [\[Kvæ95\]](#) and [\[Kam94\]](#). These variations, which form the basis of the application software, are all included in the following method:

1. Select a waveform segment of N points about the seed detection. The segment should contain a statistically significant subsegment of both noise (only) and signal (plus noise). This segment is labelled “Onset Search Window” in [Figure 7\(a\)](#). Ideally, the segment consists of two and only two statistically distinct regions about the onset time—noise only before and signal only after it.
2. Select an interval containing no signal (the “Noise Window” in [Figure 7\[a\]](#)), and from the noise characteristics construct a noise prediction filter. Select an interval consisting primarily of signal (the “Signal Window” in [Figure 7\[a\]](#)), and from the signal characteristics construct a signal prediction filter.

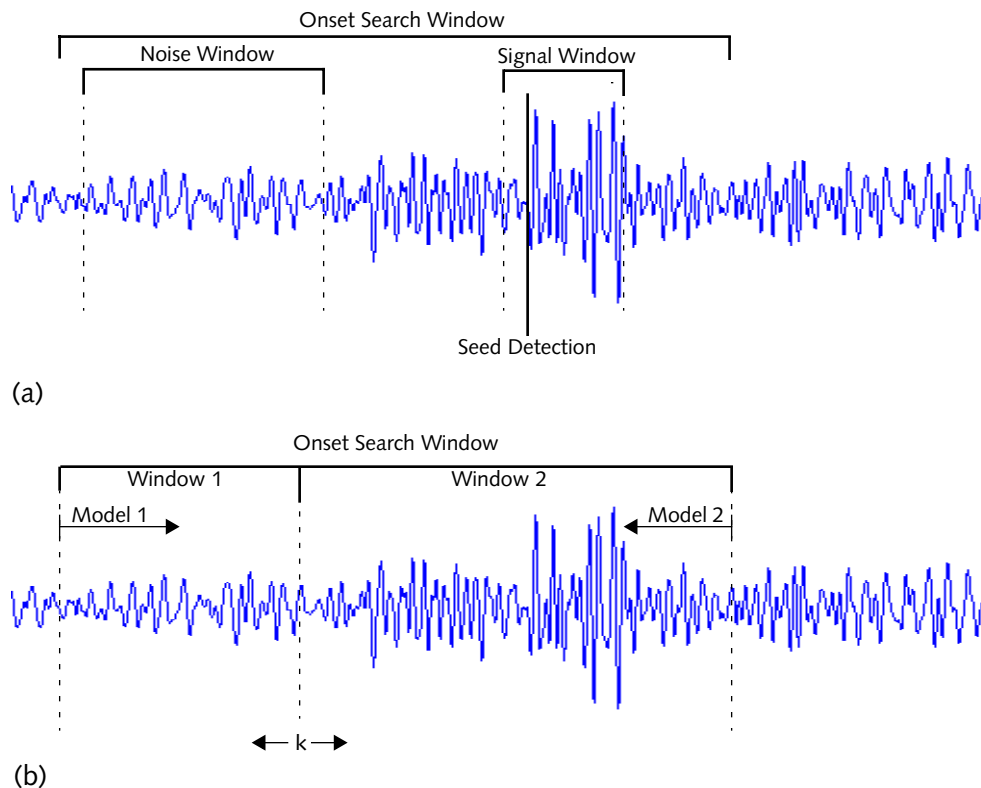


FIGURE 7. ONSET TIME ESTIMATION WINDOWS

3. Independently assign one of the prediction models to be Model 1 and one to be Model 2. Models 1 and 2 can be the same or different.

Repeat the following Steps [4](#), [5](#) and [6](#) for every sample k , where $k = 1, \dots, N$ (the pivot point) within the interval.

4. Divide the Onset Search Window into two subintervals about k , the first Window 1 and the second Window 2 as in [Figure 7\(b\)](#), with k belonging to Window 2.
5. Compute residuals e_j for Model 1 in Window 1 and Model 2 in Window 2, and estimate residual variance, σ_1^2 and σ_2^2 in each:

$$\sigma_1^2 = \sum_{j=1}^{k-1} \frac{e_j^2}{k-1}, \quad \sigma_2^2 = \sum_{j=k}^N \frac{e_j^2}{N-k+1} \quad (1)$$

6. Evaluate the AIC $\Phi(k)$ for k as follows:

$$\Phi(k) = (k-1)\log\sigma_1^2 + (N-k+1)\log\sigma_2^2 + C \quad (2)$$

7. Determine the sample k at which the AIC is minimized. This value is taken as the onset time.

This implementation of the AIC method can have many configurable variations. Options are specified through parameters that make up user onset-recipes, snr-filter-recipes and filter-recipes.

The following issues must be considered when implementing the AIC method:

- choosing the AIC prediction model
- constructing the noise and signal models
- filtering
- evaluating the models

These issues are discussed further in [“Arrival-time Estimation” on page 228](#).

Optimizing Onset Time Parameters

Arrival time parameters are optimized by comparing the automatically calculated arrival times with some set of reference arrival times and repeatedly adjusting the parameters to obtain a better statistical agreement between the two. In general, the process is performed independently for each station or even station/channel combination, because the characteristics of background noise (and to a lesser extent seismic phase energy) have a limited range of statistical properties that are specific to each site or sensor. In practice a number of different parameter combinations that span the plausible parameter ranges should be tested for each station. A typical calibration for each of the prepared parameter recipes proceeds as follows:

1. Run the automatic onset time estimator on each detection in the calibration data set for which a reference onset time has been previously determined.
2. Statistically measure the conformance between the automatically estimated arrival times and the reference arrival times. Statistics of the onset time residuals (computed time less reference time) such as mean-square or mean-absolute deviation are good candidate measures of the agreement.

The recipe that results in the best overall agreement between the automatic and reference onset times as defined in Step 2 is taken as the calibrated recipe.

Ideally, the calibration data set for a given sensor would consist of real data from the sensor, and the true onset times would be used as the reference. In reality, true onset times are available only for synthetic seismograms. Although synthetics can be effective tools for calibration, generating them to mimic the particular statistical properties of the noise signals in the real data is not straightforward. More often, real seismograms are used for the calibration data set and analyst estimated onset times proxy for the true times. The shortcoming of this approach is that onset time errors, especially systematic errors, made in human analysis carry over into the calibration results.

Measuring Amplitude and Period

Amplitude and period measurements made during station processing are used for calculating magnitudes in later processing. A variety of amplitude measurements can be made using the *DFX* application ([\[Wah96a\]](#), [\[Wah96b\]](#)), but only two are made during station processing (process 5 in [Figure 3](#)). One amplitude measurement is used for m_b magnitudes and one is used for M_L magnitudes. The results of the amplitude (*amp*) and period (*per*) measurements are stored in the **amplitude** database table for later use.

Amplitude and period measurements for m_b calculations are made on a vertical component coherent beam filtered with a three-pole, zero-phase, Butterworth filter that passes frequencies between 0.8 and 4.5 Hz. For arrays, the beam is steered to the azimuth and slowness of the detecting channel. For 3-C stations the vertical channel is used.

The waveform is screened for small excursion peaks and troughs that do not exceed a specified percentage of the maximum amplitude in the window. The amplitude is then measured as half the maximum peak-to-trough (or trough-to-peak) amplitude difference in the window (see [Figure 8](#)). This window starts 0.5 seconds prior to the automatically picked arrival time and ends 5.5 seconds after this arrival time. The period associated with the measured amplitude is calculated as twice the time between the two peaks used for the amplitude. After the period is known, the amplitude is corrected for the response of the bandpass filter and the instrument. Instrument responses are kept in files whose names are stored in the **instrument** database table. The resolutions of amplitude and period measurements are determined by the quantization and sampling rate.

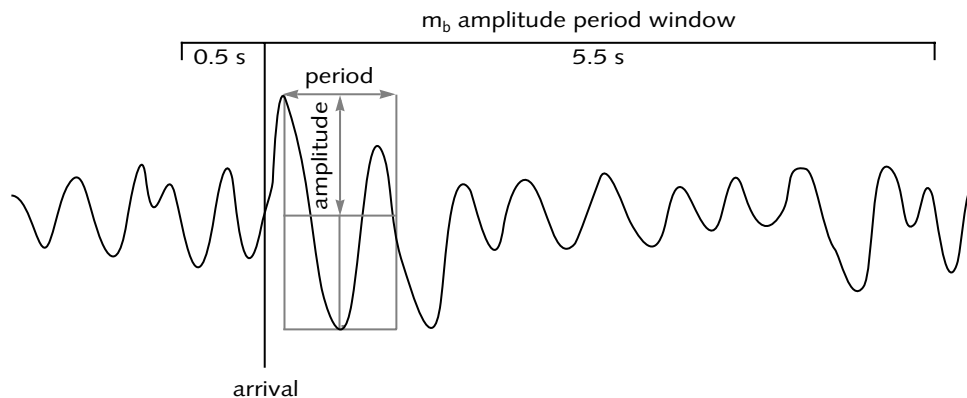


FIGURE 8. AMPLITUDE AND PERIOD ESTIMATION FOR m_b

The amplitude used for M_L calculation is measured on an incoherent beam of vertical component data filtered with a three-pole, causal, Butterworth filter that passes frequencies between 2.0 and 4.0 Hz. The beam is not steered for arrays.

The algorithm searches a four-second time window beginning at the arrival time for the maximum one-second short-term average (*stav*) amplitude. To correct for noise preceding the signal detection, a noise amplitude long-term average (*ltav*) is calculated as the running sum of *stav* values in a 30-second weighted window, normalized to a one-second equivalent length. The amplitude is corrected for noise and recorded in units of scaled counts (counts multiplied by the instrument calibration factor, *ncalib*) using the following equation:

$$amp = \sqrt{stav^2 - ltav^2} \quad (3)$$

amp is stored in the **amplitude** table of the database along with *stav* and the *snr* ($= stav/ltav$). This measurement is performed over the entire band from 2 to 4 Hz, which has a center frequency of 3 Hz and a corresponding period of 0.33 seconds. Consequently, the period is nominally referred to as 0.33.

The amplitudes listed with M_L in the event lists and bulletins are not given in scaled counts, but rather in nanometers (nm). Scaled counts are converted to nm by multiplying the scaled counts value by $A/(3*ncalper)$, where *A* is the amplitude, *ncalper* is the calibration period from the **instrument** table, and the factor 3 represents the center frequency in Hz of the bandpass for the incoherent channels.

Optimizing Amplitude and Period Parameters

Amplitude and period measurements contribute primarily to magnitude estimation, and changes in the parameters that are used to measure amplitude and period can introduce systematic bias into the magnitudes. These parameters have not been changed in recent years.

Estimating Azimuth and Slowness

Azimuth and slowness are key attributes for identifying phases, associating them with events, and estimating the location of events (process 6 in [Figure 3](#)). Frequency-wavenumber (f-k) analysis is used to estimate the azimuth and slowness of signals detected in array station data. Polarization analysis is used to estimate the azimuth and slowness of signals detected in 3-C station data.

Frequency-wavenumber Analysis

Frequency-wavenumber analysis is used on array station data to estimate the azimuth and slowness of signals. The raw array data from a short time window around the arrival time are transformed into a wide frequency band power spectrum in two-dimensional slowness space known as the f-k spectrum (see [“Frequency-wavenumber Analysis” on page 206](#)). Peaks in the f-k spectrum indicate the azimuths and slownesses of signals travelling across the array. To obtain better estimates of the azimuths and slownesses, two-dimensional interpolation is used around the peak power. The best estimates of the azimuth and slowness are written to the **detection** table of the database as *seaz* (azimuth) and *slow* (slowness) and to the **arrival** table as *azimuth* and *slow* (slowness is written in units of s/km in the **detection** table and in units of s/degree in the **arrival** table).

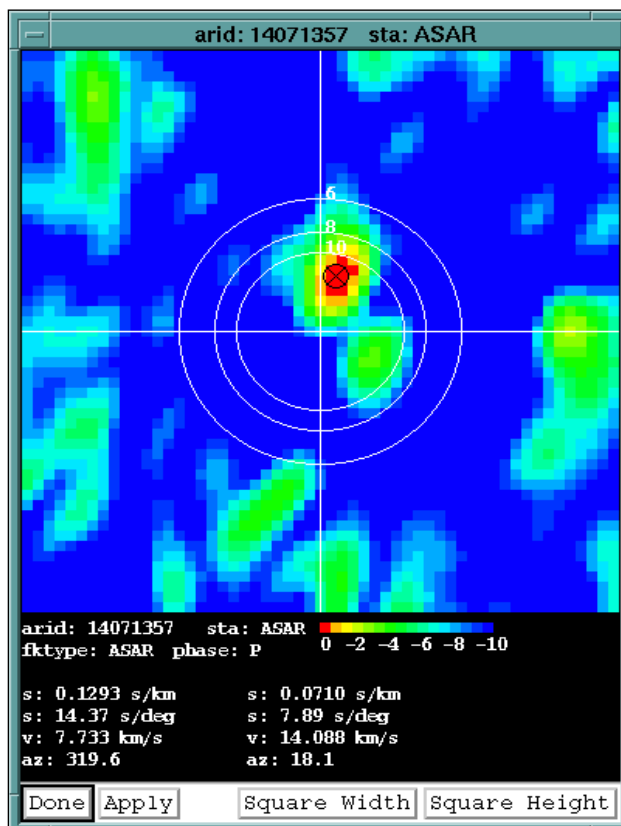


FIGURE 9. F-K ANALYSIS OF TYPICAL P ARRIVAL AT ASAR

The F-statistic, which characterizes the coherency of the f-k power spectrum, is the basis for azimuth and magnitude slowness error estimates at seismic arrays. The F-statistic is also a measure of the coherent snr at the specified phase velocity. The error estimates provided by the f-k spectrum are based upon a sum of two variances. The first variance is an estimate of the modeling error, which is input to the f-k spectrum as a recipe parameter (fk-ds); the second variance is the measurement error, which is estimated as the difference between the vector slowness coordinates of the global peak in the f-k spectrum and the vector slowness values on the contour 2 dB below the peak. This latter variance is adjusted for the snr and the center frequency of the waveform data input to the f-k spectrum and is scaled

by a second recipe parameter (fk-dk). The azimuth and slowness errors are stored in the **detection** and **arrival** tables as *delaz* and *delslo*. When station-specific slowness-azimuth corrections (SASCs) are used, the modeling error (fk-ds) is set to zero and is read from the SASC files.

Polarization Analysis

Polarization analysis is used on 3-C data to estimate the azimuth and slowness of signals. Slowness cannot be measured directly from 3-C data. Instead, the angle of incidence of the signal is estimated, and the slowness is derived using a parameterized phase velocity.

Polarization analysis is applied using the method of Jurkevics [\[Jur88\]](#) (see [“Polarization Analysis” on page 215](#)). Various features are computed from the eigenvalues of the polarization ellipsoid for a series of overlapping windows that surround the detection and are stored in the **apma** table. Azimuth (*seaz*), slowness (*slow*), and their errors (*delaz*, *delslo*) are estimated and stored in the **detection** and **arrival** table of the database for the window in which the data are maximally rectilinearly polarized.

Optimizing Azimuth and Slowness Parameters

Optimizing azimuth and slowness parameters requires a comparison of the values measured in the data by using a number of different recipes and the values that are predicted from the actual location of the event. The recipes must span a reasonable parameter space, and the data set must be reviewed by expert analysts and include locations and the theoretical slowness values for each phase.

Because slowness and azimuth are closely tied to determining signal type as well as event location, optimizing the parameters for azimuth, slowness, and signal type are combined into a single process. This process is described in [“Optimizing Signal Type Features” on page 35](#).

ASSOCIATING STATION PHASES

Associating station phases includes determining the type of signals that are detected, placing the signals into station association groups that have the characteristics expected of phases from the same event, and identifying their phase names based on the grouping (processes 7 and 8 in [Figure 3](#)).

Determining Signal Types

Signals detected in seismic data are categorized as one of four initial wave types; noise, teleseismic P, regional P, or regional S that are used for grouping detections and for phase identification (see process 7 in [Figure 3](#)). Signals are characterized based on their features and qualities extracted from the waveforms near the arrival time. Two different methods are used to determine the signal type: a neural network and simple rules.

If the station is an array, then the horizontal velocity (inverse slowness) of the signal and the quality of the f-k measurement are used to determine the signal type through simple rules. [Table 4](#) shows the velocity ranges of seismic signal types.

TABLE 4: HORIZONTAL VELOCITIES OF SEISMIC SIGNAL TYPES

Horizontal Velocity (km/s)	Signal Type
> 11	teleseismic P
5.7–11	regional P
2.9–5.7	regional S
< 2.9	noise

A detection from an array may also be classified as noise if the f-k quality is 4 and the F-statistic is less than 3 (see [“Frequency-wavenumber Analysis” on page 206](#)). Station-specific values of the velocities, the f-k quality, and the F-statistic may be set to customize the rules for each station.

If the data are from a 3-C station, then a neural network is applied to determine the signal type (see [“Neural Networks” on page 229](#) and [\[Ser93\]](#)). The neural networks use the signal period, the polarization features (described in [“Polarization Analysis” on page 215](#)), and the context of the given arrival (the relative time difference and the number of arrivals in a window centered on the arrival of interest) as inputs and produce a probability for each of the four classes as output. In practice, three neural networks are used; the first distinguishes between noise and signal, the second distinguishes between teleseismic and regional phases, and the last determines if the regional phases are P or S. Neural networks must be “trained” using the features of known signal types. If a neural network has not been trained specifically for the station being considered, an *average* neural network may be used instead (described in [“Training Seismic Neural Networks” on page 36](#)).

As an alternative to neural networks, 3-C station signal type can be determined through rules that use the frequency (*freq*), horizontal-to-vertical ratio (*hvrat*), and the rectilinearity (*rect*) features of the signal (see [Table 5](#)). These default rules do not attempt to identify noise.

TABLE 5: DEFAULT RULES FOR ASSIGNING SIGNAL TYPES

Rules	Signal Type
<i>hvrat</i> > 1.0 or <i>rect</i> < 0.7	regional S
<i>rect</i> ≥ 0.7 and <i>hvrat</i> ≤ 1.0 and <i>freq</i> ≥ 3 Hz	regional P
<i>rect</i> > 0.7 and <i>freq</i> < 3 Hz and <i>hvrat</i> ≥ 1.0	teleseismic P

The thresholds included in the rules in [Table 5](#) are the default values; station-specific thresholds can be set to customize the rules.

The last stage of determining the signal type is the possible revision of teleseismic to regional P if certain criteria are met. This step is not considered until all detections have been assigned a signal type. Teleseismic can be revised to regional P only when a compatible regional S arrival is found, and regional P is an allowed signal type. If the station is an array, then the velocity from f-k analysis must also be less than the minimum teleseismic velocity.

Optimizing Signal Type Features

Features are extracted from the detected signals by the *DFX* application (see [\[Wah96a\]](#), and [\[Wah96b\]](#)). The features are then used by the *StaPro* application [\[IDC7.1.2\]](#) to determine the signal type. The parameters that control feature extraction can be tuned to optimize the signal-type determination. Optimization consists of exploring the *StaPro* configuration parameter space for signal type as well as slowness and azimuth and selecting the *DFX* parameter sets (recipes) that result in the best classification performance. In practice, experience with data from a variety of stations makes it possible to place reasonable limits on the range of the various parameters explored. The data set used for optimization includes detections obtained from an optimized detector that have been reviewed by analysts.

The following steps can be followed to optimize feature extraction for determining signal type:

1. Rerun the post-detection processor (such as polarization or f-k analysis) off-line for each detection and for each of the standard recipes.
2. For each recipe, automatically classify each detection using, for example, classical discriminant analysis [\[Sut91\]](#) or neural networks for large enough data sets.
3. For each recipe and for the appropriate subset of detections, compare the measured azimuth and slowness with those predicted from a reference location (this step applies only to Primary seismic stations).

4. Select the recipe with the highest performance for classification or the highest combined performance for classification and for azimuth and slowness estimation.

Training Seismic Neural Networks

A three-stage neural network is used for determining the initial wave type of arrivals recorded at 3-C seismic stations. The training determines the weights of the neural network nodes for a training set. Nonspecific neural network weights are used for many stations, but specific weights can be determined for stations that have enough data to train the networks.

Training the neural networks must accommodate the imperfect knowledge of the noise because detections that were not associated to events during analysis may be signals from events that do not meet the event definition criteria. An adaptive training method is used as follows [\[Wan97\]](#):

1. Select random teleseismic phases that have not required the onset time correction by analysts for the training set.
2. Select regional P and S phases that have onset corrections less than two seconds for the training set. This correction range is less restrictive than the one for the teleseismic phases so that sufficient data can be obtained.
3. Select random noise phases that are not associated to events in the bulletins for the training set. This step requires caution. Although some of these phases are likely to be true noise, many are signals from small events that were too small to include in the bulletin. Ideally, an expert analyst should identify the signals to be classified as noise.
4. Train the stage of the neural net that separates noise from signal using the entire training set.
5. Because some signal is included in the noise training set, remove the noise detections misclassified as signal from the training set and train the neural network again.

6. As the next step, all of the remaining training data that were not correctly classified are removed from the training set and the neural network is trained again.
7. The remaining stages of the neural network are trained using standard methods because the noise signal types have been removed during earlier stages, and the remaining training data are “known” through analysis.

Optimizing Signal Type Rules

Optimizing the rules for signal type determination consists of comparing, in order, the results of applying a specific rule to a “control” data set to the outcome that should have resulted given the true nature of the signal. A suite of reasonable “thresholds” for each rule are tested, and the threshold that maximizes the percentage of signals that are classified correctly and minimizes the number that are classified incorrectly is chosen. This choice should be made with an understanding of the impact on the final bulletin quality.

Grouping Signals and Identifying Phases

After all detections at a station have been assigned a signal type, they are collected into disjoint *station association groups*, within which each signal has the characteristics expected of phases from the same event (process 8 in [Figure 3](#)). The first arrival in each group is called the generator of the group, and its signal type is a local P, regional P, or a teleseismic phase. Subsequent “test” arrivals in each group are identified on the basis of compatibility with the generator, as determined by the arrival times and compatibility tests (described in the next section).

Three categories of the distance from the station to the event are considered: teleseismic, regional, and local. Events at teleseismic and local distances are treated as simple cases in station processing, and the arrivals from these events are omitted prior to the more general regional analysis.

The general approach to grouping is to proceed in time order, identifying the next occurring generator and all compatible arrivals in a specified time window. As the arrivals are grouped, the phases are identified to be compatible with the grouping.

Testing for Compatibility

At array stations, azimuth compatibility is established when the azimuth difference between the generator and test arrival is less than or equal to the product of a constant factor and the root mean square (rms) of the standard deviation of the two azimuths. The default factor is 2.0 for teleseismic generators, 2.5 for local generators, and 3.0 for regional generators.

Amplitude compatibility at an array station occurs when the amplitude ratio of the test arrival and generator are within a specified range. The default range is from 0.025 to 40.

Both azimuth and amplitude compatibility tests must be satisfied, otherwise the test arrival is dropped from consideration of membership in the group.

A different compatibility test is applied if azimuth data are not available or are considered unreliable (for example, 3-C station S arrival). For these cases, a more tightly constrained amplitude test with the default range is from 0.4 to 25.0 is applied. Additionally, the test arrival must be within 120 seconds following the generator for the test arrival to be compatible.

Grouping Phases from Teleseismic Events

Arrivals with teleseismic signal type (T) are grouped first. Proceeding in time order, the first teleseismic phase not already belonging to a group is taken as the next generator of a teleseismic group. Any subsequent teleseismic signal within a specified arrival time interval (default is 90 seconds) after the generator is assigned to the same group if the test arrival is compatible with the generator (as previously defined). The generator is assigned a phase identification of "P," and the subse-

quent arrivals in the teleseismic group are assigned an initial phase identification of "tx," indicating that they are probably detections in the coda of the P. The group of teleseismic signals are then removed from subsequent phase grouping analysis.

Grouping Phases from Local Events

Special rules are used to group phases of events close to a station. If an arrival with signal type of P is closely followed by a compatible S, the pair is considered to have come from a local event, and the arrivals form the basis of a station association group. Specifically, a P-S pair is identified as a local event whenever the S arrival time is within a specified time interval (25 seconds) after the P generator, and the S is compatible with the generator (as previously defined). To avoid erroneously taking a late phase in the P coda of a regional event as the generator of a local event, the rules require that a regional P phase must not closely precede (25 seconds) the candidate generator with which the candidate generator is compatible. For local events, the P generator is given the phase identification "Pg." The S phase is assigned "Lg" unless its frequency is low (less than 1.54 Hz) and, it is close to the Pg (less than 65 seconds), implying that the event is close, in which case it is assigned "Rg." The local group is then completed by adding any P coda phase, which might be present as compatible P phases between the Pg and the S (but not past 20 seconds from the Pg). The P coda phases are given the phase identification "Px."

Grouping Phases from Regional Events

After noise detections and arrivals associated with teleseismic events and local events have been identified and grouped, the remaining arrivals are analyzed under the general rules for regional events. As before, the first remaining non-grouped P signal type is taken as the generator of a new regional group. Intervening non-grouped S signal types preceding the next generator are not included in any group; they are given the phase identification "Sx." The earliest compatible S signal type after the generator (within 360 seconds) is paired with the regional P generator to form the basis of a new group. The group is then completed by

including any additional compatible regional S phases arriving within a specified multiple (2.1) of the time between the generator and the first regional S and any compatible regional P phases between the generator and the first regional S (but not beyond 20 seconds from the generator).

Either the first or the largest S-type detection in the group is identified by Bayesian analysis as explained in the next paragraph. The remaining detections in the group are identified by travel-time prediction. If no compatible regional S signal type is close enough to the generator to be included in the group, the group is either degenerate (containing the generator only), or it consists of the generator and some P coda phases. In either case, the generator is identified as "Pn" or "Pg" based on an intermediate velocity threshold of 7.5 km/s. Associated P coda phases are labelled "Px."

A Bayesian analysis technique is used to identify either the first S in a regional group or the largest one based on a static table of empirically-determined conditional probabilities (see ["Bayesian Inference" on page 232](#) and [\[Bac93\]](#)). The Bayesian analysis estimates probabilities for all possible S-type regional phases (Sn, Lg, Rg, and Sx). The regional phase with the highest probability is the most-likely candidate, but additional constraints are also applied. This technique is more robust for Sn than the other regional phases, and it is most-likely that Sn is not part of a group if Sn is not chosen during Bayesian analysis. The conditional probabilities are based on velocity, horizontal-to-vertical power ratio, S-P time, period, and context with respect to other S phases. The context categories are only-S, first-S-of-two, first-S-of-many, or largest-S. The static table can include the conditional probabilities for multiple stations, and a default table can be specified for stations that have not been calibrated.

The prediction procedure begins by estimating a preliminary event location using the arrival time and azimuths of the generator and the S-type arrival whose phase identification was determined by Bayesian analysis. The arrival time of one of the remaining regional phases (Pg, Lg, Rg) is then predicted. Three conditions must be met for a phase to be predicted: 1) the phase cannot already exist in the group, 2) the group must have an arrival with a compatible initial wave-type, and 3) the event depth and distance must be consistent with constraints for the phase being

considered. If an arrival that is consistent with the prediction is found, then its phase is identified, and it is used to refine the preliminary event location. The procedure continues until all valid regional phases have been considered.

LOCATING AND ESTIMATING MAGNITUDES OF SINGLE-STATION EVENTS

The final task in seismic station processing is to compute a single-station event location and estimate local magnitude for association groups that satisfy a user-specified event definition criteria (process 9 in [Figure 3](#)). The initial event locations and magnitudes are used for phase association during network processing to screen detections from small local events.

Event confirmation is based on a weighted-count of defining observations (arrival time, azimuth, and slowness). The count uses individual weights for each type of defining observation; separate weights are available for primary arrival times (1.0), secondary arrival times (0.7), array azimuths and slownesses (0.5), and 3-C azimuths and slownesses (0.25). An association group passes the event confirmation test if its weighted count exceeds a user-defined threshold (2.6). Event locations are determined for these groups (see ["Event Location" on page 119](#)), but the event location process may reduce the number of defining attributes. Reducing the number of defining attributes causes a reduction of the weighted count, so the weighted count test is applied again. Finally, an additional confirmation test based on the size of the error ellipse is applied.

Local magnitudes are estimated for confirmed events by using the method described in ["Estimating Local Magnitudes" on page 237](#). The local network magnitude is computed as a weighted average over individual station magnitudes (developed from regional P or Pn). The magnitudes are based on the short-term-average amplitude measured on a 2-4 Hz incoherent beam (arrays) or vertical channel (3-C stations).

Hydroacoustic Station Processing

This chapter describes how the IDC processes data from the hydroacoustic IMS stations and includes the following topics:

- [Overview](#)
- [Checking Data Quality](#)
- [Detecting Signals](#)
- [Extracting Signal Characteristics](#)
- [Associating Station Phases](#)

Hydroacoustic Station Processing

OVERVIEW

Station processing includes the following tasks:

- checking the quality of the data
- filtering data
- detecting signals
- extracting time features
- extracting energy features
- extracting moment features
- extracting cepstral features
- extracting miscellaneous features
- determining signal types
- grouping and identifying the phases of the signals at each station

These tasks are performed automatically on small time increments of data (10 minutes) as they arrive using the *DFX* and *StaPro* applications (see [\[Wah96a\]](#), [\[Wah96b\]](#), and [\[IDC7.1.2\]](#)). *DFX* performs the quality check, generates channels for detection processing, detects signals in the data, and extracts the signal features from the waveforms. *StaPro* uses these extracted features to determine the signal type, which in turn gets used in network processing (see [“Network Processing” on page 73](#)). The results of station processing are reported in the SLSD.

Two types of hydroacoustic stations exist in the IMS: hydrophone and T phase stations. Hydrophone stations consist of one or more hydrophones located offshore of an island at a depth corresponding to the [SOFAR](#) channel axis (this is the depth

at which sound propagates most efficiently in the oceans). T phase stations are 1-component (1-C) seismic stations located near shore, typically on mid-ocean islands. The difference in the station processing between the two types of stations is primarily a function of the different sampling rates.

The significant station processing steps are shown in [Figure 10](#) and are described in the following sections. [Table 6](#) lists the software application that performs each of the station processing steps.

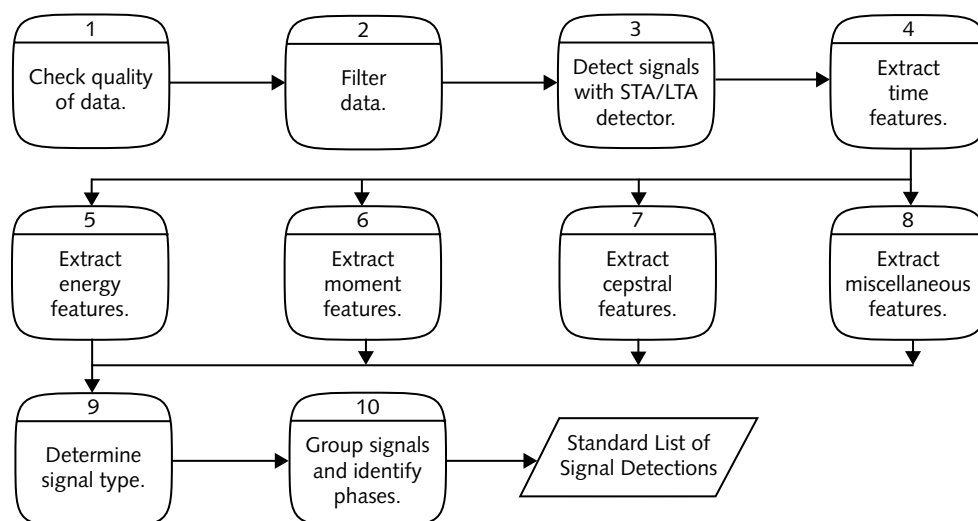


FIGURE 10. STATION PROCESSING OF HYDROACOUSTIC DATA

TABLE 6: SOFTWARE USED IN STATION PROCESSING

Step	Software	Reference(s)
1 Check quality.	DFX	[Wah96a] , [Wah96b]
2 Filter data.		
3 Detect signals.		
4 Extract time features.		

Step	Software	Reference(s)
5 Extract energy features.		
6 Extract moment features.		
7 Extract cepstral features.		
8 Extract miscellaneous features.		
9 Determine signal type.	<i>StaPro</i>	[IDC7.1.2]
10 Group and identify phases.		

Checking the quality of the data is the first step in station processing (process 1 in [Figure 10](#)). The process is the same for seismic, hydroacoustic, and infrasonic data (see “[Checking Data Quality](#)” on page 10).

Detecting signals in hydroacoustic data includes frequency filtering the data and then processing the filtered data with an STA/LTA detector (processes 2 and 3 in [Figure 10](#)). Two different phases are sought: “H” phases that originate from in-water events and “T” phases that originate from terrestrial events. [Figure 11](#) shows examples of a T and an H phase (the time scale differs between the two examples). Both T and H phases can vary widely in shape, and these examples do not represent T and H phases in general. Although a hydroacoustic station can have multiple sensors, traditional array processing techniques cannot be used for the large sensor separation. Instead the sensors are treated as individual stations.

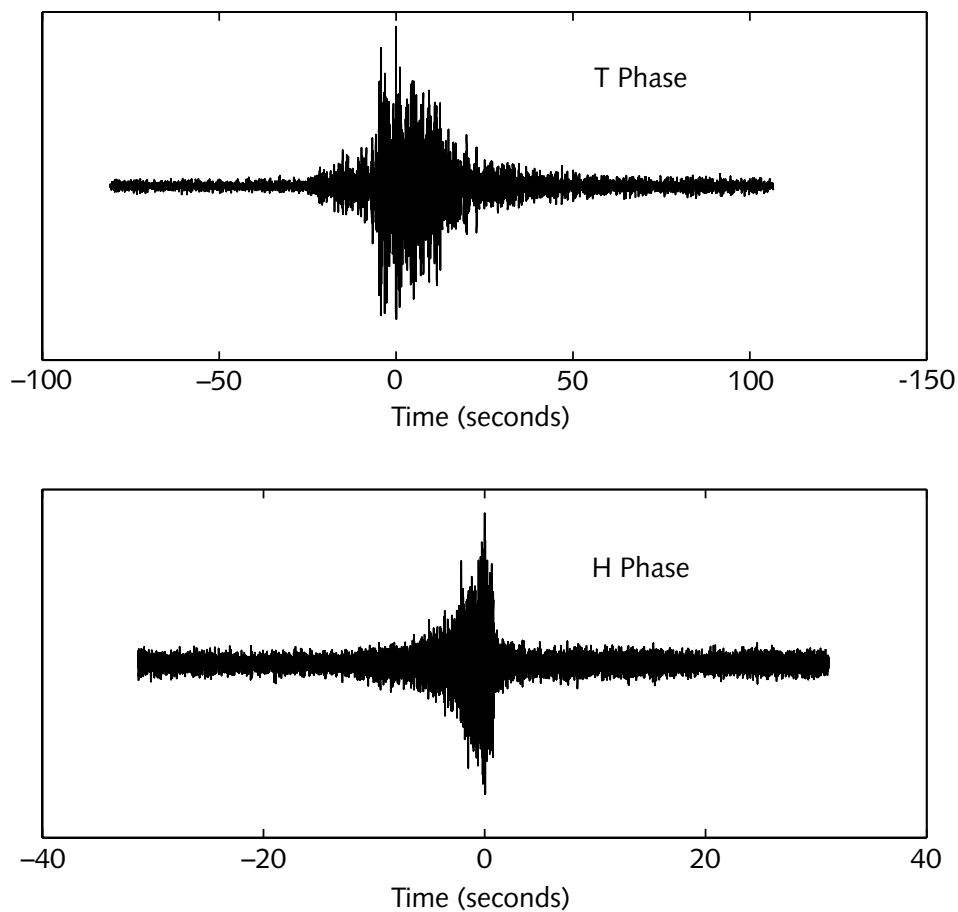


FIGURE 11. EXAMPLES OF T AND H PHASES

Filtering

The filter bands used for detecting signals are selected to enhance the signals of interest while attempting to avoid noise (process 2 in [Figure 10](#)). Because noise characteristics can vary among stations, the filter bands and their sensitivity may be set accordingly.

detector. An arrival is declared when a detection is on any of the channels (however, if desired, this convention may be changed to require detections on two or more channels to declare an arrival).

EXTRACTING SIGNAL CHARACTERISTICS

After a signal has been detected, an instrument response correction filter (if available) is applied to the data. The filter is designed to correct the time series amplitude to units of pressure (μPa). Next, features of the signal are extracted from the data for each channel (processes 4 through 8 in [Figure 10](#)). Time features must be extracted first, but after the time features are known, the remaining features may be extracted in any order. The various features measured are used to identify the arrival type, associate the arrival to an event, locate the event, and characterize the event. The *DFX* application attempts to extract features from each of the channels even if a particular filter band has no detection (see [\[Wah96a\]](#), [\[Wah96b\]](#)). The extracted features, listed in [Table 8](#), are stored in the database in the **hydro_features** table. The onset and termination times of the signal are the first features to be estimated. This process is attempted for each filter band and if successful, the remaining features listed in [Table 8](#) are computed from the channel data. The probability-weighted time (defined in the following sections) for the channel with the highest total energy is used as the arrival time of the detection and is entered in the **arrival** table. The features from all the frequency bands are used in *StaPro* to determine the signal type of the detection.

TABLE 8: FEATURES EXTRACTED FROM HYDROACOUSTIC DATA

Features	Hydro_Features Attribute
onset and termination times	<i>onset_time, termination_time</i>
probability-weighted time and uncertainty	<i>prob_weight_time, sigma_time</i>
signal summation time	<i>total_time</i>
total energy	<i>total_energy</i>

50

Termination time is set when the snr drops below a given threshold, which is usually larger than the threshold used for onset time (see [a] in [Figure 12](#)). To prevent early termination, the time series is searched for this threshold only after two conditions have been met. The first condition requires that prior to termination the snr reaches a given threshold (b), which is generally some multiple of the detection threshold. If this condition cannot be met within a preset interval (c), no features are measured in this frequency band. The second condition requires that the termination occurs after the maximum snr in another preset interval (d). In addition to the previous two conditions, the signal length is not allowed to exceed a set maximum (c) even if the termination time if the snr has not dropped below the termination threshold.

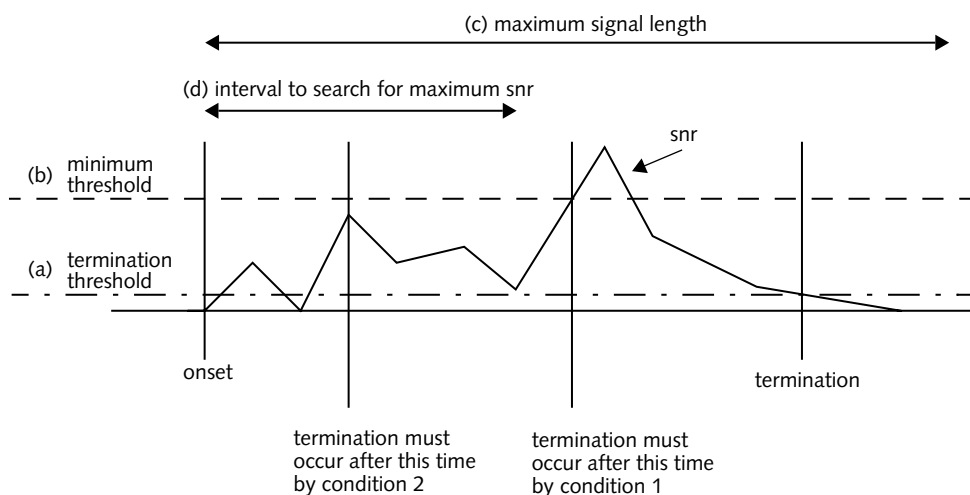


FIGURE 12. TERMINATION TIME ESTIMATION

Probability-weighted Time and Uncertainty

Consistently measured arrival times are necessary for locating events. Because T phases are emergent and have long durations, a consistent estimate of arrival time cannot be made using onset times or peak amplitudes. Instead, the probability-weighted time is used as the arrival time estimate for hydroacoustic phases (see [“Probability Weighted Time” on page 227](#)).

The probability-weighted time is an estimation of the peak energy arrival taking into account the uncertainties in a measured value being the “real” peak. The uncertainties in the measured values are estimated from the pre-event noise. Thus if the signal has a high snr and one “large” peak is present, the probability weighted time will occur at that peak arrival time. When several similar amplitude peaks occur in the wave packet, the probability-weighted time occurs between the peaks, similar to a weighted average. In addition to estimating the arrival time, the probability-weighted time procedure calculates an error estimate for the arrival time.

Signal Summation Time

Signal summation time (*total_time*) is the amount of time during which the snr is above a specified threshold. This measure is useful to distinguish a long continuous signal from a long duration signal consisting only of spikes (which is often an indication of noise).

Estimating Energy Features

Energy features of hydroacoustic data (process 5 in [Figure 10](#)) include the total energy, the intensity average time, and the peak energy and time.

Total Energy

The total energy is used first to define the channel on which the arrival time is estimated and later to identify the signal type. Total energy is estimated by calculating the sum of the signal squared minus the average noise:

$$total_energy = \sum_{i=1}^N (x_i^2 - ave_noise) \Delta t \quad (4)$$

where x_i is the i th sample of the data between the onset and termination times. The average noise level (*ave_noise*) is calculated for a time window (by default 15 seconds long) prior to the detection:

$$ave_noise = \frac{\sum_{i=1}^{N_{noise}} x_i^2}{N_{noise}} \quad (5)$$

Both *total_energy* and *ave_noise* are converted to dB relative to 1 μ Pa prior to being recorded in the database.

Intensity Average Time

The intensity average time (*mean_arrival_time*) is the average time between onset and termination times weighted by the energy in the signal at each sample, x_i .

$$mean_arrival_time = \frac{\sum_i (x_i^2 - ave_noise) \cdot t_i}{total_energy} \quad (6)$$

Mean_arrival_time is used as a neural network input to help determine the signal type.

Peak Energy and Time

The peak time (*peak_time*) is the time between onset and termination times for which the largest absolute signal level occurs. Peak energy (*peak_level*) is the square of the maximum signal level recorded in dB relative to 1 μ Pa. The *peak_time* and *peak_level* are included in the neural network inputs to help determine the signal type.

Estimating Moment Features

Moment features of hydroacoustic data (process 6 in [Figure 10](#)) are the time spread, skewness, and kurtosis.

Signal Time Spread, Skewness, and Kurtosis

The signal time spread (*time_spread*), *skewness*, and *kurtosis* are included in the parameters to determine the signal type. They correspond to the second, third, and fourth moment statistics, respectively. The *time_spread* is an estimate of the duration over which the energy in the signal is spread. The *skewness* is a measure of the lopsidedness of the distribution of energy over time. *Kurtosis* is a measure of how compressed the signal is in time. The signal whose energy envelope function approaches a Gaussian will have a 0 *kurtosis* by the definition in equation 9. An envelope with larger tails than a Gaussian will have a positive *kurtosis*, and an envelope that approaches a delta function will have a value of -3 . The formulae to calculate the moment statistics are as follows:

$$time_spread = \sqrt{\sum_i \frac{(x_i^2 - ave_noise)(t_i - mean_arrival_time)^2}{total_energy}} \quad (7)$$

$$skewness = \frac{1}{time_spread^3} \cdot \sum_i \frac{(x_i^2 - ave_noise)(t_i - mean_arrival_time)^3}{total_energy} \quad (8)$$

$$kurtosis = \frac{1}{time_spread^4} \cdot \sum_i \frac{(x_i^2 - ave_noise)(t_i - mean_arrival_time)^4}{total_energy} - 3 \quad (9)$$

Estimating Cepstral Features

Cepstral analysis is used to detect periodicities in the power spectrum of the signal (process 7 in [Figure 10](#)). The power spectrum of a signal that contains two or more coherent arrivals will exhibit strong periodic oscillations even if the separate arrivals are not apparent in the time domain. The secondary arrivals may be associated with the reverberations caused by the expansion and collapse of the bubble

generated by an explosion (bubble pulse) but can also be due to reflections or volcanic tremors. The periodicity in the spectrum of an in-water explosion is shown in [Figure 13](#).

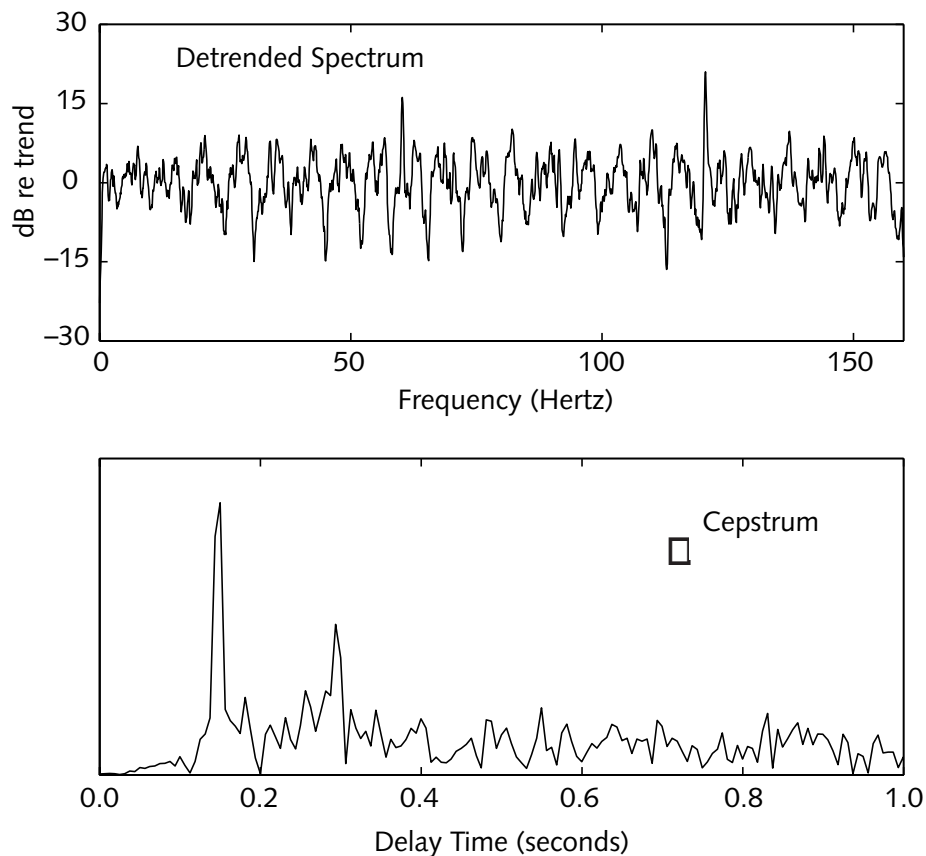


FIGURE 13. DETRENDED SPECTRUM AND CEPSTRUM FOR H PHASE

A cepstrum is the Fourier transform of the logarithmically scaled (in magnitude) power spectrum of the signal. If a periodicity exists in the power spectrum, the cepstrum will display a prominent peak corresponding to a time delay between arrivals in the time series. An example is shown in [Figure 13](#) at a delay time of 0.15 seconds.

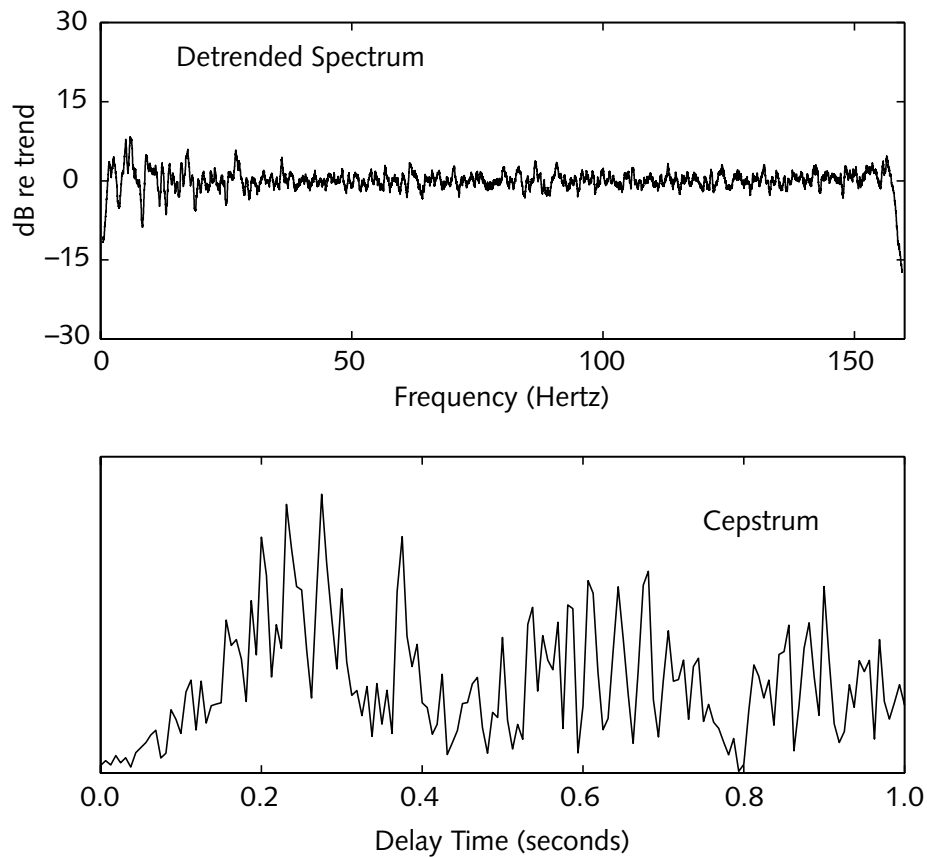


FIGURE 14. DETRENDED SPECTRUM AND CEPSTRUM FOR T PHASE

For comparison, [Figure 14](#) shows the detrended spectrum and the corresponding cepstrum for a T phase. No obvious periodicity is in the spectrum, and consequently no particular peak in the cepstrum is obvious.

Cepstral analysis is most usefully applied to broadband data (2–80 Hz). The cepstrum is computed, and a set of parameters that characterize the signal is extracted. The delay time between pulses is estimated from the largest peak in the cepstrum between some minimum and maximum time delay (*cep_delay_time_signal*). The size of this peak is measured relative to the variance in the cepstrum,

which provides a measure of the significance of the cepstral peak—(see [“Algorithms” on page 175](#)).

Estimating Miscellaneous Features

Currently only one miscellaneous feature, the number of threshold crossings, is extracted from hydroacoustic data (process 8 in [Figure 10](#)).

Number of Threshold Crossings

The number of threshold crossings (*num_cross*) used to help determine the signal type is a count of the number of times the pressure squared signal crosses the onset threshold between the onset and termination times.

ASSOCIATING STATION PHASES

Associating station signals includes determining the type of signals that are detected, placing them into station association groups within which each signal has the characteristics expected of phases from the same event, and identifying their phase names based on the grouping (processes 9 and 10 in [Figure 10](#)).

Determining Signal Type

The features extracted from each channel are used to assign a signal type to the detection (process 9 in [Figure 10](#)). Hydroacoustic signal types include T (terrestrial), H (water), and N (noise). The signal type is used to group detections and associate phases to events.

The station processing system uses either a neural network (see [“Neural Networks” on page 229](#)) or a set of explicit rules applied to the extracted signal features (see [Table 8](#)) to determine the signal type. A neural network uses a set of weights to map the signal features to a signal type. The weights are derived by “training” the neural network from a station specific data set that has previously been analyzed by hand. For a standard reference on neural networks see [\[Bis95\]](#).

Feature	Definition
energy ratio	$total_energy$ (32–64 Hz) - $total_energy$ (3–6 Hz) (dB)
duration	$termination_time$ - $onset_time$ in 3–6 Hz band
time spread	$time_spread$ in 3–6 Hz band
fractional time	$total_time$ (3–6 Hz) / duration
crossing density	num_cross (3–6 Hz) / duration

1. Does the phase have
 - energy ratio < -15.5 (dB), and
 - crossing density > 12.0 , and
 - time spread > 5 seconds?

2. If the phase does not satisfy Rule 1, does the phase have
 - energy ratio < -15.5 (dB), or
 - fractional time < 0.4 , or
 - crossing density > 20.0 , or
 - duration < 6.0 seconds, or
 - time spread > 35.0 seconds?

March 1999 IDC-5.2.1

3. Are extracted features that would preclude a T or N identification in the previous rules missing?

If so, then identify the arrival as a N phase.

4. If all the parameters are defined and do not match any of the previous criteria, then identify the arrival as an H phase.

The numeric values used in the previous rules can be modified to improve performance.

Optimizing Signal Type Features

The parameters that control feature extraction are tuned to optimize the signal type determination. Optimization consists of exploring the configuration parameter space for signal type and selecting the parameter sets (recipes) that result in the best classification performance. In practice, experience with data from a variety of stations makes it possible to place reasonable limits on the range of the various parameters explored. The data set used for optimization includes detections obtained from an optimized detector that have been reviewed by analysts.

The following steps are performed to optimize feature extraction for determining signal type:

1. Rerun the post-detection processor offline for each detection and for each of the standard recipes.
2. For each recipe, automatically classify each detection using, for example, classical discriminant analysis [\[Sut91\]](#), or for large enough data sets use neural networks.
3. Select the recipe with the highest performance for classification.

Training Hydroacoustic Neural Networks

Neural networks are trained by using large sets of reviewed data for which the signal type is known (in general, several hundred sets are used for each phase type). Few H phases are actually recorded so synthetic waveforms must be used to aug-

ment the training sets. The training determines the weights of the neural network nodes for a particular station. The training sets must use the features extracted in automatic processing and may not use analyst added or altered arrivals.

Training is accomplished in two stages. First a neural network is trained to separate T from H or N phases. Then a second neural network is trained to differentiate between H or N phases. This differentiation is necessary because of the significant overlap of parameters between the phase types (particularly H and N).

Rule Optimization

Optimizing the rules for signal type determination consists of comparing, in order, the results of applying a specific rule to a “control” data set, to the outcome that should have resulted given the true nature of the signal. A suite of reasonable thresholds for each rule are tested, and the threshold that maximizes the percentage of signals that are classified correctly and minimizes the number that are classified incorrectly is chosen.

Grouping Signals and Identifying Phases

Hydroacoustic signal grouping (process 10 in [Figure 10](#)) is based on signal duration overlap and signal type. Few groupings occur due to the longer duration of individual detections. A time ordered search for the next non-associated T arrival determines a group generator. A subsequent arrival may be associated with the generator to form a group if it has a signal type similar to the generator and its signal duration is within that duration estimated for the generator. Phase identification for hydroacoustic arrivals consists of evaluating the relative positions of the station associated signals. The signal having the largest amplitude (in the **arrival** table) is the primary arrival for that group and is identified as a T or H phase, depending on the signal type. Other arrivals in the group are identified as secondary arrivals (Tx or Hx).

When the phase has been identified, the signal characteristics are incorporated into the SLSD.

Infrasonic Station Processing

This chapter describes how the IDC processes station data from the infrasonic IMS stations and includes the following topics:

- [Overview](#)
- [Checking Data Quality](#)
- [Detecting Signals](#)
- [Extracting Signal Characteristics](#)
- [Generating Maximum Coherence Trace](#)
- [Associating Station Phases](#)

Infrasonic Station Processing

OVERVIEW

Station processing includes the following tasks:

- checking the quality of the data
- identifying periods of spatial coherence
- detecting signals
- estimating arrival time
- estimating azimuth and slowness
- generating a maximum spatial coherence trace
- grouping and identifying phases

These tasks are performed automatically on small time increments of data (10 minutes) as they arrive using the *DFX* and *StaPro* applications (see [\[Wah96a\]](#), [\[Wah96b\]](#), and [\[IDC7.1.2\]](#)). The significant station processing steps are shown in [Figure 15](#) and described in the following sections. [Table 10](#) lists the software application that performs each of the station processing steps. The results of station processing are placed in the SLSD.

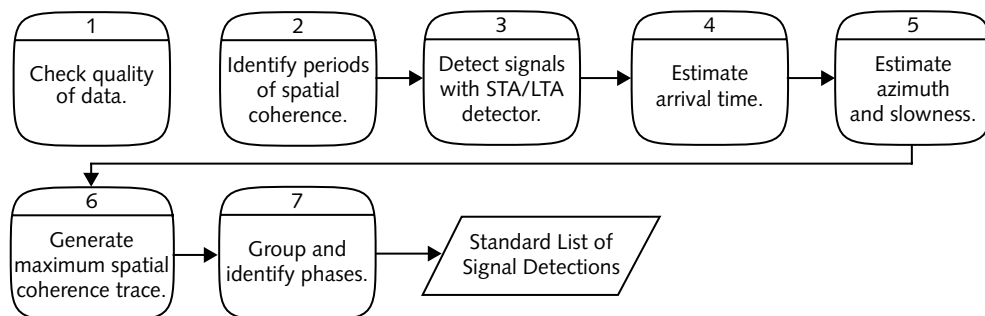


FIGURE 15. STATION PROCESSING OF INFRASONIC DATA

TABLE 10: SOFTWARE USED IN INFRASONIC STATION PROCESSING

Step	Software	Reference(s)
1 Check quality.	DFX	[Wah96a] , [Wah96b]
2 Identify periods of spatial coherence.		
3 Detect signals with STA/LTA detector.		
4 Estimate arrival time.		
5 Estimate azimuth and slowness.		
6 Generate maximum spatial coherence trace.	StaPro	[IDC7.1.2]
7 Group and identify phases.		

CHECKING DATA QUALITY

Checking the quality of the data is the first step in station processing (process 1 in [Figure 15](#)). The process is the same for seismic, hydroacoustic, and infrasonic data (see [“Checking Data Quality” on page 10](#)).

DETECTING SIGNALS

As with seismic and hydroacoustic signal detection, the fundamental basis for deciding that a detection ought to be declared is the sudden presence of excess waveform energy in one or more spectral bands of waveform data. Infrasonic array data exhibit strong ambient noise (from wind bursts) that is incoherent between individual sensor sites within an array station. The small numbers of sensor elements in typical infrasonic arrays lead to broad beam response patterns, and only moderate snr gains are realized through forming beams. Consequently, energy detection alone, applied to beamed bandpass filtered waveform data, yields far too many false detections. To overcome this problem, the infrasonic detection algorithm exploits the inherent spatial coherence of airborne infrasonic arrivals from distant sources versus the spatial incoherence of the ambient noise field. Signals may be detected (using an STA/LTA detector) only during periods of high spatial coherence.

Identifying Periods of Spatial Coherence

Periods of spatially coherent energy are identified before attempting to detect signals (process 2 in [Figure 15](#)). Values representing the spatial coherence of the energy are calculated for points in two-dimensional slowness space. For each time interval, “slowness planes” are generated for several different combinations of frequency bands and sensor channel subsets. Detection threshold, spectral band, and the sensor channels subset may be set on a point-by-point basis in the slowness plane. For normal infrasonic detection processing these latter three parameters are generally kept at fixed values within a given slowness plane.

To calculate the spatial coherence for a slowness plane, data are filtered using Finite Impulse Response (FIR) bandpass filters into the frequency bands desired (see [Table 11](#) and the algorithm description in [“Finite Impulse Response Filters” on page 185](#)). Normalized cross-correlation functions are then calculated for all of the processing intervals for all of the spectral bands and for all of the independent data channel pairs required (see [“Normalized Cross-correlation Functions” on page 193](#)).

TABLE 11: TYPICAL FILTERS USED IN INFRASONIC SIGNAL PROCESSING

Geometric Mean Interstation Separation (km) at the Sensor-Array Site (Typical Site)	Low Band Band Pass (Hz)		Medium Band Band Pass (Hz)		High Band Band Pass (Hz)	
0.15 (PDIAR, SGAR)	0.70	2.80	1.00	3.80	1.40	6.40
0.16 (LSAR)	0.40	1.60	0.60	2.30	0.80	3.20
0.76 (DLIAR)	0.12	0.50	0.24	1.00	0.50	2.00
1.3 (TXIAR)	0.06	0.24	0.12	0.50	0.24	1.00
0.51 (WRAI – short baseline group)	0.08	0.32	0.16	0.60	0.32	1.20
3.2 (WRAI – large aperture group)	0.02	0.08	0.04	0.16	0.08	0.32

A time succession of slowness planes is generated using the normalized cross-correlation functions for each specific combination of frequency band and sensor channel subset. For a given slowness plane, the cross-correlation functions to be used are determined by the sensor channels, which have been designated in the beam-group (within the file of beam-recipes). For example, if the beam-group specifies that only sensors 1, 3 and 4 are to be used, then for this beam-group the independent sensor pairs are (1,3), (1,4), and (3,4); normalized cross-correlation functions are combined for these sensor pairs only. At each point in the slowness plane, the coordinates of the point (azimuth, magnitude slowness) determine the time delays (lags) at which the contributing cross-correlation functions are sampled. The spatial coherence value at each point in the slowness plane is the average of the cross-correlation values each evaluated at the appropriate lag for that sensor pair.

Detecting spatial coherence is the process of detecting peaks in the slowness planes. The slowness planes are searched for spatial coherence levels in excess of thresholds (set in the beam-recipe parameter files). Each point whose calculated

coherence level is within a fixed fraction of the threshold level and whose coherence level is larger than its four nearest neighboring points is considered a potential coherence peak. Where two or more potential peaks are close to each other in slowness, only the point with the largest coherence level is considered for further processing, thereby reducing redundant detections. The slowness coordinates of each potential coherence peak are used to calculate a refined slowness plane, which covers a limited region of slowness space with a fine gridwork of newly calculated coherence values. If the largest coherence value among the points in this refined slowness plane exceeds the threshold, then the point is declared a “contact,” and the azimuth and slowness of the contact are estimated. Where one or more contacts are identified in a given slowness plane, some ancillary attributes of the best such contact are estimated and retained.

The contacts in the successive slowness planes are used to define the time(s), spectral band(s), and sensor channels wherein spatially coherent energy is present. The slowness coordinates of the contacts define the directions of arrival and slownesses of the potential detections.

For a given time interval of processing, typically 10 to 60 minutes, contacts are grouped into “clusters” regardless of time, spectral band, or subset of sensor channels used. Where one or more clusters are formed, these clusters of contacts are required to be distinct in slowness space, either in azimuth, slowness, or both coordinates. Within a given cluster, the spectral band, the sensor channels, and the slowness coordinates of the contact with the highest level of spatial coherence are attributed to the cluster.

Detecting

For each cluster that is formed, the appropriate bandpass filtered waveform data are beamed to the azimuth and slowness attributed to the cluster to form a time series, $B(t)$. An STA/LTA energy detector is applied to $B(t)$ (see [“Infrasonic STA/LTA Detectors” on page 201](#)). In addition, a time series, $F(t)$, of exact F-statistic values (see [“F-statistics” on page 196](#)) is calculated for a sliding time window using the

vector slowness coordinates attributed to the cluster as well as the same bandpass filtered waveform data and the same sensor channels as were used to generate $B(t)$ (process 3 in [Figure 15](#)).

An infrasonic signal is detected when the STA/LTA detector applied to $B(t)$ exceeds its threshold, and the F-statistic time series $F(t)$ also exceeds its threshold. A new infrasonic signal can be detected only after the outputs of the STA/LTA detector applied to $B(t)$ and/or the F-statistic time series has fallen below the threshold state.

This process is continued until all of the waveform data for a given cluster is fully evaluated for detections. Each cluster is processed in this same fashion. If the data have more than one detection, and the times of the detections are within a specified range for a given cluster, then the detection with the lower snr is eliminated. The detections that remain are written simultaneously to both the **arrival** and the **detection** tables of the database.

Optimizing Detection Processing of Infrasonic Data

In general, many more signals are detected by the automated system than contribute in some way to an event location or origin time. Therefore, it is important to minimize the number of signals that do not contribute to events while still detecting enough phases to capture and locate all events. This goal is accomplished by adjusting the input processing parameters for each station according to the conditions that are experienced at that station and requiring that the features extracted from each signal satisfy criteria set to favor the infrasonic signals from explosive sources.

The key processing parameters, which govern the performance of the infrasonic detection algorithm, are as follows:

- bandpass limits of the spectral bands
- coherent integration time
- STA and LTA averaging times, as well as the time gap between the STA and LTA windows

- threshold value to be applied to the coherent detector
- threshold value to be applied to the STA/LTA detector
- extent and sampling strategy of the search space in vector slowness

EXTRACTING SIGNAL CHARACTERISTICS

Estimating Arrival Time

The time of the first sample for which the STA/LTA detector applied to $B(t)$ exceeds its threshold and the F-statistic time series $F(t)$ also exceeds its threshold is used to estimate the initial time of a declared detection (process 4 in [Figure 15](#)). The initial time is used to define the start-time of the waveform data segment, which is evaluated to estimate spectral features of the detection.

The arrival time of an infrasonic detection is estimated as the time of peak STA level, in the interval of time immediately following the initial time for which $B(t)$ was above its detection threshold. The error of the measurement (*deltim* in the **arrival** and **detection** tables) is estimated from a Cramer-Rao calculation, which uses an estimate of the signal bandwidth, the noise bandwidth, the coherent integration time, and the coherent-signal-to-incoherent-noise ratio.

Estimating Azimuth and Slowness

Arrival azimuth and magnitude slowness are estimated using an internal invocation of the traditional f-k spectrum calculation. The waveform data input to this f-k spectrum calculation are pre-aligned in time to place the detection at nominally zero vector slowness in the slowness plane. The f-k spectrum is calculated for only a small range of highly refined vector slowness values. The net estimate of the detection's azimuth and slowness is the vector sum of the "pre-alignment" vector slowness and the "vernier" estimate of vector slowness from the "pre-aligned" f-k spectrum. The variance in each of these parameters is estimated from the normal f-k spectrum calculation of azimuth and slowness variances adjusted for the "pre-alignment" vector slowness value at which the f-k spectrum was "centered."

When the f-k spectrum cannot be calculated, an infrequent occurrence that is generally caused by marginal waveform data quality in the vicinity of a detection, then a modified Cramer-Rao bound is substituted for variances normally calculated by the f-k spectrum. This Cramer-Rao estimation is based upon the site geometry, the snr of the refined slowness plane with the highest spatial coherence level, and the spectral density function of the bandpass filtered and beamed data (see [“Modified Cramer-Rao Variance Estimation” on page 212](#)).

GENERATING MAXIMUM COHERENCE TRACE

For each processing time interval, the best spatial coherence level exhibited during that time interval regardless of spectral band, data channels, or slowness plane coordinates are determined (process 6 in [Figure 15](#)). This best spatial coherence level is set in a time series ($t, v(t)$) that can be displayed during analysis alongside the waveform data to indicate time intervals of high spatial coherence.

ASSOCIATING STATION PHASES

Associating infrasonic phases includes determining the signal type to be either signal or noise, placing the signals into association groups; and identifying their phase names based on the grouping (process 7 in [Figure 15](#)).

Determining Signal Type

In the context of airborne infrasonic signals received at distances in excess of a few tens of kilometers, one to three principal and distinct transport paths exist due to the atmospheric temperature variation with altitude. The temperature variation with altitude causes a corresponding sound speed (magnitude group velocity) variation with altitude. Where this sound speed profile exhibits minima, upward incident sound energy is refracted downward towards the earth's surface where it may “bounce” upward again, subject to some absorptive losses. Such sound speed minima define the transport paths or “channels.” Prevailing winds at altitude, as well as the season, govern the shape of the sound speed profile. The

**Infrasonic ▼
Station
Processing**

sound channel at 95 to 130 km altitude is known as the *Thermospheric* channel, the channel at approximately 18 to 50 km altitude is known as the *Stratospheric* channel, and the channel below 20 km is known as the *Tropospheric* channel. At this time no attempt is made to distinguish between Tropospheric, Stratospheric, and Thermospheric arrivals. Additional knowledge is needed before arrivals can be further broken down. Instead, infrasonic signals and noise are separated by a velocity threshold. The horizontal velocity range used for an infrasonic signal is from 270 to 660 m/s. Figure 16 shows a short duration airborne infrasonic impulsive signal. The first arrival phase travelled via the Stratosphere whereas the short duration second arrival traveled through the Thermosphere.

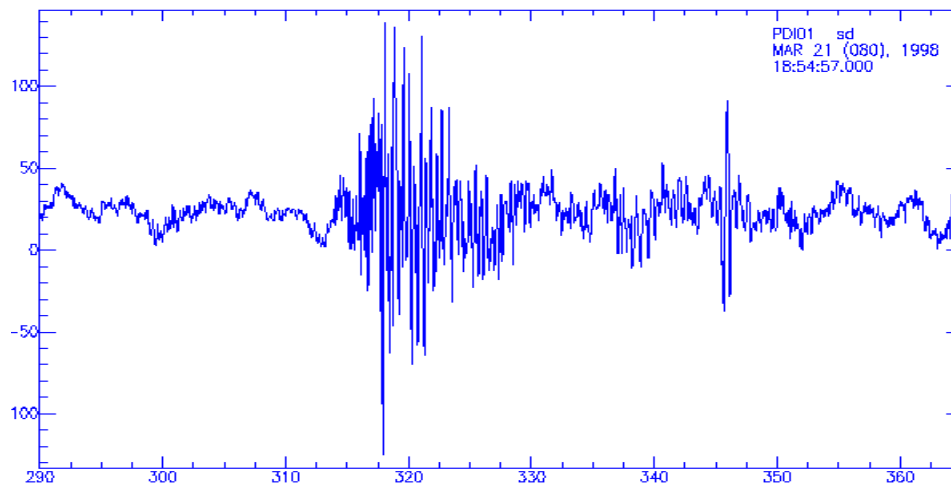


FIGURE 16. STRATOSPHERIC AND THERMOSPHERIC ARRIVALS

Grouping Signals and Identifying Phases

Infrasonic signals are grouped similarly to seismic data (process 7 in [Figure 15](#)). Infrasonic signal phases eventually will be labelled in accordance with the physical path of the sound channel responsible for energy transport of that phase, as in Thermospheric phase or Stratospheric phase. Without sufficient experience to reliably identify phases to this degree in automatic processing, initial phase identification will label the first signal in a group as "I" and each subsequent signal in that group as "Ix." A time-ordered search for the next non-associated I arrival determines a group generator. The azimuth compatibility test determines if subsequent non-associated arrivals of similar signal type and occurring within a user-controlled time window following the generator are compatible. The default time window is station specific and ranges from 300 to 600 seconds. All other signals in the group are labeled Ix.

Azimuth compatibility occurs when the azimuth difference between the generator and test arrival is less than or equal to the product of a constant factor and the rms of the standard deviation of the two azimuths. The default factor is 2 for infrasonic generators.

When the phases have been identified, the signal characteristics are incorporated into the SLSD.

Network Processing

This chapter describes how the IDC processes signals detected on a network of sensors for the purpose of locating and characterizing events. It includes the following topics:

- [Overview](#)
- [Developing Grids](#)
- [Associating Phases Based on Grid Points](#)
- [Extracting Large Events](#)
- [Analyzing Splits and Redundancy](#)
- [Analyzing Location and Outliers](#)
- [Resolving Conflicts](#)
- [Predicting Phases](#)
- [Checking Event Consistency](#)
- [Tagging Arrivals for Subsequent Processing](#)
- [Time Windowing and Time Stepping](#)
- [Retrieving Auxiliary Seismic Station Data](#)

Network Processing

OVERVIEW

The primary purpose of network processing is to associate those arrivals detected and identified during station processing with events and to locate and estimate the size or magnitude of the events. [Figure 17](#) shows the main steps in network processing. The *Global Association (GA)* application performs all of the tasks of network processing except for requesting auxiliary seismic data (see [\[LeB96a\]](#), and [\[LeB96b\]](#)). For efficiency, network processing is applied to several source regions and time intervals separately (processes 1 through 6 in [Figure 17](#)). When all of the source regions and time intervals have been processed, conflicts between time intervals and regions are resolved and final quality checks are made (processes 8 through 11 in [Figure 17](#)). The *Retrieve Subsystem* performs the request function (process 12 in [Figure 17](#)). Standard Event Lists SEL1, SEL2, and SEL3 are the products of network processing. [Table 12](#) lists the software application that performs each of the network processing steps.

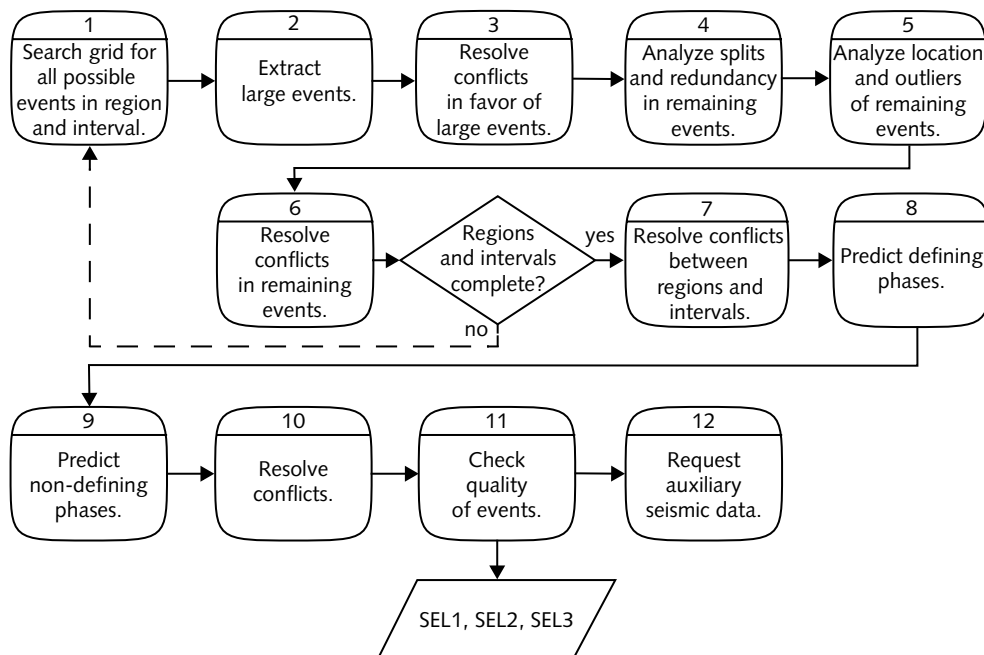


FIGURE 17. NETWORK PROCESSING MODEL

TABLE 12: SOFTWARE USED IN NETWORK PROCESSING

Step	Software	Reference(s)
1 Search grid for events.	GAassoc	[LeB96a] , [LeB96b]
2 Extract large events.		
3 Resolve event conflicts.		
4 Analyze splits and redundancy.		
5 Analyze location and outliers.		
6 Resolve event conflicts.		

TABLE 12: SOFTWARE USED IN NETWORK PROCESSING (CONTINUED)

Step	Software	Reference(s)
7 Resolve conflicts between regions.	<i>GAconflict</i>	[LeB96a] , [LeB96b]
8 Predict defining phases.		
9 Predict nondefining phases.		
10 Resolve event conflicts.		
11 Check event consistency.		
12 Request auxiliary seismic data.	<i>WaveExpert</i>	

The phases identified during station processing are associated with events using a grid-based event search procedure (GA) that is computationally efficient (and predictable) given the large number of stations in the IMS networks (see process 1 in [Figure 17](#), [\[LeB96a\]](#), and [\[LeB96b\]](#)). The approach is similar to generalized beam forming [\[Rin89\]](#), but with additional information and logic applied to resolve the many conflicts that arise.

In a grid-based approach, a grid of cells covers all of the surface of the earth plus depth zones at which S/H/I events are known historically to occur. Each grid cell is searched for events by examining all arrivals at the stations nearest the grid cell. Arrivals at these stations are called *driver* arrivals. The *driver* is used to predict the time of a hypothetical event at the grid cell, and this constitutes a seed event. The event origin time is used, in turn, to predict the arrival times for phases at stations other than the *driver* station. If a phase is predicted at the observed time for a detection at another station and is compatible with the *driver*, then the predicted arrival is added to the seed event. If several stations have phases compatible with the *driver*, the event has a higher probability of being real. A preliminary event hypothesis is confirmed when it reaches a critical size and passes prelocation event-definition criteria. The results of this process provide an exhaustive list of where and when events might have occurred and the phases that might be associated with these events.

The presence of a large event (detected at many stations) in a data set typically generates numerous preliminary event hypotheses clustering in space and time around the location and origin time of the actual event. Large event hypotheses are extracted and analyzed before all others to reduce the processing load and eliminate conflicts efficiently near the start of the processing sequence (processes 2 and 3 in [Figure 17](#)).

The large set of event hypotheses next is refined by testing the consistency of the associated phases with the locations of the events and resolving conflicts between phase associations (processes 4, 5, and 6 in [Figure 17](#)).

After the individual time intervals and source regions have been analyzed (processes 1 through 6 in [Figure 17](#)), the combined results are processed to resolve conflicts between the events in the individual time intervals and source regions, to associate remaining phases, and to check the consistency of the events prior to publication as event lists (processes 7 through 11 in [Figure 17](#)).

The network processing functions shown in [Figure 17](#) are performed a number of times prior to expert review, first to obtain early results from data that arrive relatively quickly and later to incorporate data on phases for which the propagation times are longer and additional data that was requested from auxiliary stations. The first pipeline uses data from only the primary seismic and hydroacoustic stations. Preliminary estimates of event locations based on the first pass of the network processing are reported in SEL1 and used to request data from auxiliary seismic stations (process12 in [Figure 17](#)). The data obtained from the auxiliary stations (after station processing) are used, along with the primary seismic, hydroacoustic, and infrasonic data, to improve the event location and magnitude estimates. These results are reported in SEL2. One final network processing pass is made just prior to human analysis to incorporate any late-arriving data. These results are reported in SEL3.

The following sections describe the algorithms used for network processing. Some of these algorithms are used many times in the sequence of steps depicted in [Figure 17](#), but they are described only once.

DEVELOPING GRIDS

To search the grid, a grid that includes the locations of the grid points and their relationships to the stations of the station networks must be developed. Two grid files are generated that contain this information: one of the grid files, which contains purely geographical information is referred to as the *auxiliary grid file*. The other grid file, which contains both geographical and propagation information from the grid points to the stations in the network, is referred to as the *grid file*. Both files are nearly static and are regenerated only when new information on phase propagation becomes available or necessary (as when a new station is added to one of the networks).

The grid used in the GA application contains both surface and depth cells (see [\[LeB96a\]](#) and [\[LeB96b\]](#)). The surface cells cover the entire surface of the earth and extend vertically into the atmosphere to include atmospheric events; the depth cells exist only where deep seismicity occurs. The information associated with each grid cell includes the stations with potential to record first arrival, distances and azimuths to the stations, phase names, travel-times, and slowness information for each phase that might be observed, and magnitude information, such as the minimum magnitude detectable at the station.

Surface Cells

Grid points are established on the earth with a quasi-uniform spacing. The grid points are distributed along equidistant parallels, including the equator, with points both at the North and South Poles.

The grid is built by starting with uniformly distributed points on the equator with one point at longitude = 0. On the next latitude line, a point is placed halfway in longitude between two points at the equator. This process is repeated for each latitude line, where one point is placed exactly halfway between two points on the previous latitude. All points on a parallel are evenly spaced. At lower latitudes, this results in a quasi-quincunx pattern rather than a square grid.

The radius of each surface cell is set to the latitude spacing. This setup provides a sufficiently large overlap to ensure complete coverage of the earth and approximates uniformity. [Figure 18](#) shows a portion of a global three-degree grid over Northern Europe.

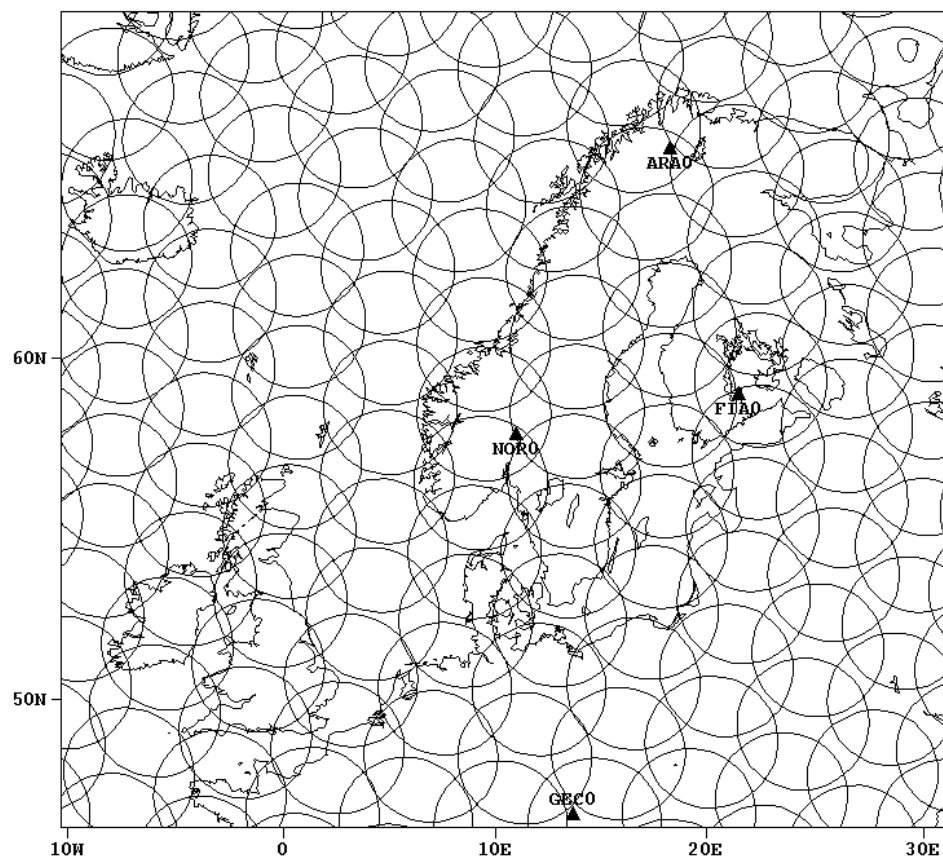


FIGURE 18. THREE-DEGREE GRID COVERAGE OVER NORTHERN EUROPE

Depth Cells

A file containing events recorded over many years from a published global seismic bulletin (such as the *United States Geological Survey Preliminary Determination of Epicenters*) guides the establishment of grid cells at depth. A set of depths and widths for the depth cells are provided by the user. Every surface location is included in the grid, and the depth cells below the surface locations are included if they satisfy a user-specified threshold on the event density from the seismicity bulletin. The cell shape is a portion of a cone truncated at the lower and upper depth of the cell. [Figure 19](#) shows some of the deep cells (centered at 130 km) in the Western Pacific. One of the cells is highlighted, and the great circle paths to potential first-arrival stations are shown for that cell.

Grid Contents

The grid file and auxiliary grid file both contain geographical information about the grid points. This information consists of the latitude, longitude, depth of the grid point, lower depth bound, upper depth bound, radius of the cell, and seismicity [b value](#) within the cell for each point. Data in both grid files are organized on a cell-by-cell basis, with all information for a cell grouped in a cell record. This data arrangement allows for natural parallelization of the processing. In addition, the grid file contains attributes that are unique to *grid point–station* pairs and others that are unique to *grid point–station–phase* triplets.

For *grid point–station* pairs, the file contains the distance between the grid point and station, the azimuth from the station to the grid point, the back-azimuth from the grid point to the station, the m_b magnitude distance-dependent correction coefficient with its distance, and depth derivatives.

For *grid point–station–phase* triplets, the information in the grid cell consists of the travel time between the grid point and the station, the minimum and maximum travel time for that phase within the cell, the travel time derivatives with respect to distance and depth, and the maximum slowness difference within the cell. The range of definition for each phase is specified within an input parameter file.

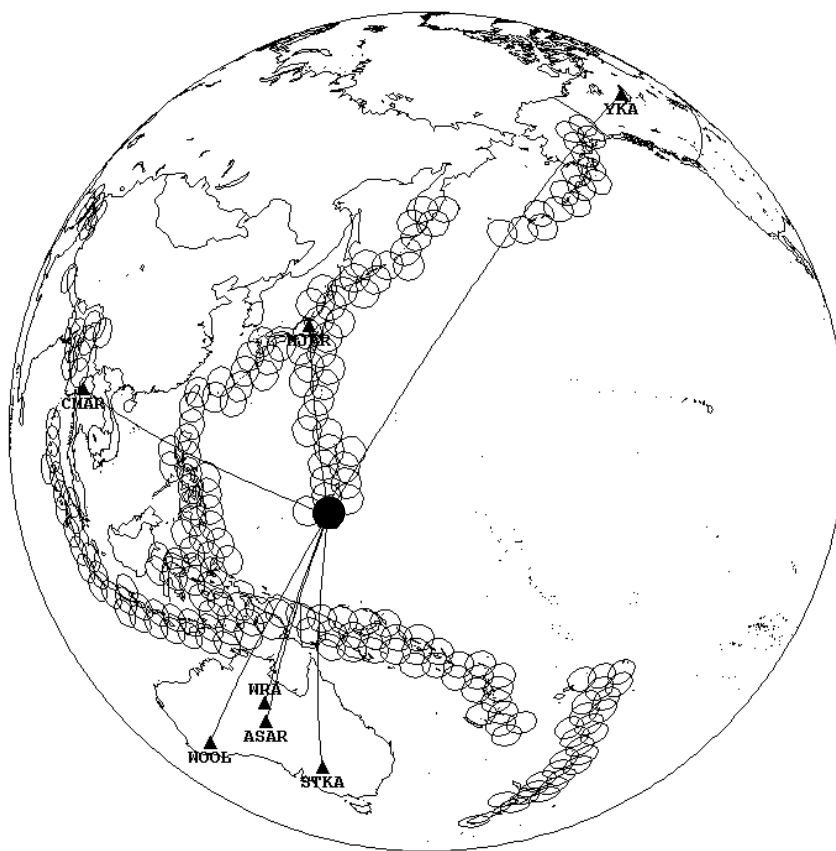


FIGURE 19. MAP OF WESTERN PACIFIC DEPTH CELLS FOR GA GRID

Acoustic Phases

In the special case of acoustic phases, the information contained within the grid is based on slightly different rules than for seismic phases.

For hydroacoustic phases, a blockage rule applies. Cells with a nonblocked path to the hydroacoustic station contain the travel time information. All hydroacoustic stations are systematically examined as potential *driver* stations. The depth range of the hydroacoustic phase is such that only surface cells will contain information about the cells.

For infrasonic phases, the travel time information is included exclusively for surface cells. The depth range of the surface cell in that case includes the entire atmosphere.

ASSOCIATING PHASES BASED ON GRID POINTS

The static grid is exhaustively searched for events and the arrivals detected in the waveform data from the stations (process 1 in [Figure 17](#)).

Identifying Driver Arrivals

Detection data are read from the **arrival**, **amplitude**, and **apma** database tables for all stations belonging to the network. All detections are loaded except for the following:

- detections that were associated to an event during station processing and were determined to be of magnitude less than a minimum detection threshold
- detections determined to be noise or with an unknown phase identification
- detections that either have been specifically requested through the data request mechanism to refine a previously constructed event hypotheses (tagged as `requested` in the **ga_tag** table) or have been previously associated during an earlier round of processing (tagged as `locked_association` in the **ga_tag** table)

After the detections have been read, *driver* arrivals for each grid cell are identified. A *driver* is an arrival at one of a limited set of stations in the network that could be expected to record the earliest arrival for an event in the given grid cell. The stations are called the *driver* stations for that cell. The maximum number of distance-ordered *driver* stations that will be considered for any grid cell can be specified by the user. In addition to belonging to one of the *driver* stations, the initial phase

identification of the *driver* detection must be one of a specified list of phases. The current list includes seismic phases P, Pn, and Pg, hydroacoustic phase H, and infrasonic phase I.

The *driver* phase must also satisfy constraints on its slowness vector. The constraint on the slowness vector is illustrated in [Figure 20](#), which shows the relationship between theoretical and observed slowness vectors in slowness (S_x - S_y) space. The length of the difference between the observed slowness vector and the predicted slowness vector must be less than a threshold. This threshold is the sum of the uncertainty of the slowness for the arrival multiplied by a specified factor (σ) and the maximum difference between the slowness vector to the center of the cell and to any other points in the cell ($\Delta \text{ cell}$).

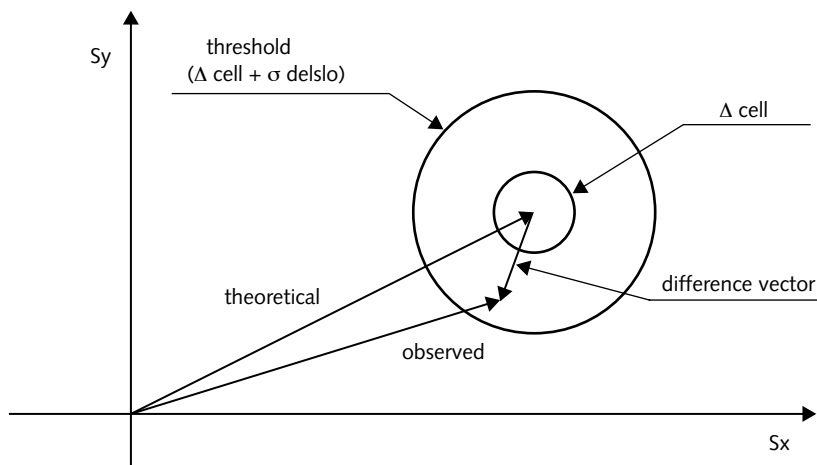


FIGURE 20. THEORETICAL AND OBSERVED SLOWNESS VECTORS FOR GA DRIVER

The standard deviation of the slowness for the arrival (*del slo*) is read from the **arrival** database table. The maximum slowness vector difference ($\Delta \text{ cell}$) for the specific station, phase, and cell is obtained from the grid file.

The slowness vector for seismic phases constrains both the distance and the azimuth between cell and station, because seismic phases exhibit simple slowness variations with distance. The slowness variations of infrasonic phases with distance, however, are complex due to multipathing within the atmosphere and are not currently modeled in detail. Therefore, the slowness does not constrain distance for infrasonic phases. The azimuth, however, is measured with high precision for infrasonic phases, and the direction of the slowness vector is necessary for restricting the search for infrasonic events. Because no component of slowness is measurable on a single hydrophone, an azimuthal constraint for hydroacoustic phases can only come from the joint processing of groups of hydrophones from which directional information can be derived.

Searching for Corroborating Arrivals

The next step in the association process is to search for arrivals that are consistent with the *driver* arrival. These arrivals are called corroborating arrivals. The search is made by examining arrivals at all stations at times later than the *driver*, where the arrivals are restricted to phases in a user-specified list. Because phases may be misidentified during station processing, specific alternative phase names to those determined during station processing for these arrivals are considered. If the confidence of the initial phase identification as read from the *belief* value in the **assoc** database table is greater than a user-specified threshold, then the phase identification determined during station processing will not be overridden. The *belief* field, however, is not currently used. An optional restriction that a secondary phase may be associated only if there is an associated primary phase from the same station can be imposed. Another optional restriction is that arrivals grouped during station processing may not be associated if the group includes a regional S, and the distance estimated from the S – P time is incompatible with the distance from the station to the grid point. Hydroacoustic phase blockage information is contained in the grid so that blocked phases can not be associated.

A simple test of the travel time and slowness vector (if available) consistency is applied to each potential corroborating arrival. An expected arrival time is estimated from the travel time for the phase stored in the grid file and an estimate of

the event origin time derived from the *driver* assuming the location is at the center of the cell. The travel-time residual must be less than the sum of the time uncertainty multiplied by a user-specified factor and the maximum travel time across the grid cell (stored in the grid file). The time uncertainty is the root mean square (rms) of *deltim* in the **arrival** database table and a modeling error, which is distance dependent. The slowness vector test is the same as that applied to the *driver* previously described.

A more rigorous chi-square statistical test is applied if an arrival successfully passes the travel-time and slowness window tests previously described. The chi-square test examines the compatibility between the corroborating arrival and the *driver* arrival with an hypothesized event within the current grid cell. The test uses all available features of the arrivals, including travel time, slowness vector, and amplitude. [Figure 21](#) illustrates the geometry of the two-station location used in the test. The use of the amplitudes is the main departure from the standard location procedures used in the *libloc* library. A detailed description of this specific chi-square test is given in [“Automatic Association Chi-Squared Test” on page 247](#).

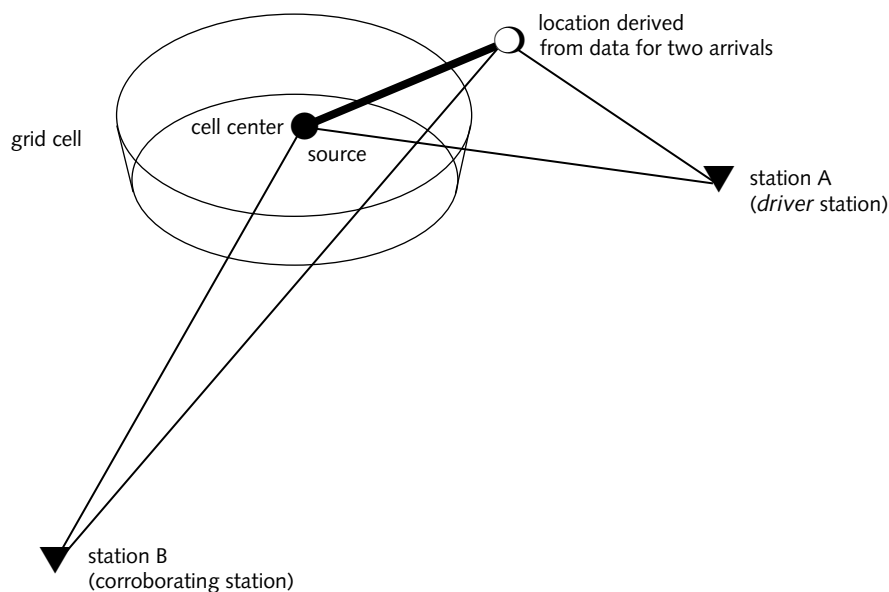


FIGURE 21. TWO-STATION LOCATION IN GA ASSOCIATION

Applying Event Definition Criteria

When all possible corroborating arrivals have been examined for compatibility with the *driver*, a set of event-screening tests are applied to the preliminary event hypothesis. If the event hypothesis does not pass the event-screening tests, it is abandoned and excluded from subsequent processing.

Weighted-Count Test

The sum of the weights (shown in [Table 13](#)) for the defining associated arrivals must be greater than 3.55 to pass the test.

TABLE 13: PHASE WEIGHTS FOR EVENT DEFINITION IN SEL1, SEL2, AND SEL3

Phase Type	Station Type	Arrival Time	Azimuth	Slowness
primary seismic	seismic array	1.0	0.4	0.4
secondary seismic	seismic array	0.7	0.4	0.4
primary seismic	seismic 3-C	1.0	0.2	0.2
secondary seismic	seismic 3-C	0.7	0.0	0.0
H	hydroacoustic	1.2	0.0	0.0
I	infrasonic	1.0	0.4	0.4

Seismic Arrival Quality Test

The arrival quality test involves the individual detection quality as measured by the *delslo* parameter and the distance between the event and station. This test is applied to screen the many accidental and fictitious hypotheses that can be formed from a small number of arrivals at teleseismic distances and normal measurement errors in the slowness vector. In the f-k space, the *delslo* parameter can be thought of as the width of the incoming beam at the station. The larger this width, the wider the intersection of this beam with the surface of the earth at the event location. The intersection of the beam with the surface is proportional to the

square of the *de/slo* parameter ($\delta\theta$). The other parameter involved in the measure of the intersection of the beam with the earth surface is the distance between the event and the station. The intersection area is proportional to $x = \sin(\Delta/2) \delta\theta^2$, where Δ is the distance between event and station, and $\delta\theta$ is the width of the beam. [Figure 22](#) illustrates the geometry of the beam of width $\delta\theta$ intersecting with the surface of the earth at a distance of Δ degrees for straight rays. The surface area of the intersection is calculated as follows:

$$\begin{aligned} A &= 2\pi R^2 \frac{(\sin(\Delta/2)\delta\theta)^2}{\cos((\pi - \Delta)/2)} \\ &= 2\pi R^2 \sin(\Delta/2)\delta\theta^2 \end{aligned} \quad (10)$$

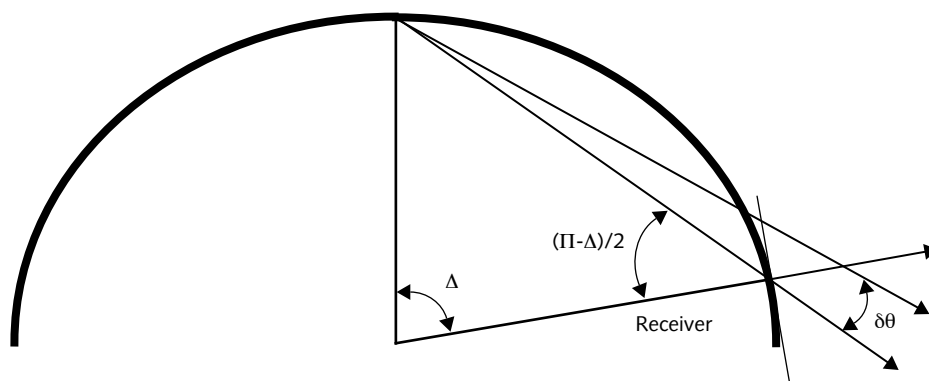


FIGURE 22. INTERSECTION OF BEAM WITH SURFACE OF EARTH

The x quantity varies between 0 and a large positive number. A rational function is used to normalize the quantity between a minimum of β and a maximum of 1. In addition, to account for the velocity variations of seismic phases, a distance-dependence has been built into the arrival quality (Q) expression:

If $\delta\theta = 0$, $Q(\delta\theta, \Delta) = \beta$

If $\delta\theta > 0$ and $\Delta < 90^\circ$:

Network ▼
Processing

$$Q(\delta\theta, \Delta) = \beta + \frac{1 - \beta}{1 + \alpha(\delta\theta^2 \cdot \sin(\Delta))} \quad (11)$$

If $\delta\theta > 0$ and $90^\circ < \Delta < 100^\circ$:

$$Q(\delta\theta, \Delta) = \beta + \frac{1 - \beta}{1 + \alpha(\delta\theta^2)} \quad (12)$$

If $\delta\theta > 0$ and $142^\circ < \Delta < 148^\circ$:

$$Q(\delta\theta, \Delta) = \beta + \frac{1 - \beta}{1 + \alpha(0.5\delta\theta^2)} \quad (13)$$

elsewhere:

$$Q(\delta\theta, \Delta) = \beta + \frac{1 - \beta}{1 + \alpha\left(\delta\theta^2 \cdot \left(3 - \frac{(\Delta - 100)}{80}\right)\right)} \quad (14)$$

The values used for parameters α and β are respectively 0.75 and 0.1 for seismic phases.

Hydroacoustic Arrival Quality Test

This test is intended to be moot for hydroacoustic phases because they do not allow the calculation of the slowness vector, and the values are set to allow three-station events to pass the test.

The values used for parameters α , β are respectively 0.0 and 0.46 for seismic phases. Because $\delta\theta = 0$ for hydroacoustic phases, $Q(\delta\theta, \Delta) = \beta$.

Infrasonic Arrival Quality Test

At all distances:

$$Q(\delta\theta, \Delta) = \beta + \frac{1 - \beta}{1 + \alpha(\delta\theta^2 \cdot \sin(\Delta))} \quad (15)$$

The values used for parameters α , β are respectively 0.35 and 0.0008 for infrasonic phases.

Application of Arrival Quality Test

The arrival quality test is applied to every event. The sum of all arrival qualities for the event is computed and compared to a threshold value. This threshold is the same for all events independent of the technologies involved and currently set at 1.2. The value of 1.2 for the threshold, 0.35 for α , and 0.0008 for β have been found to optimally eliminate seismic false alarms while retaining events that analysts have historically identified as real.

At this stage in the processing of preliminary events, the distance Δ used is the distance between the station and the center of the grid cell. The same test is applied at several places in the processing sequence. After location, an exact distance between an event and a station is available, and that exact distance is used.

EXTRACTING LARGE EVENTS

The presence of a large event (detected at many stations) in a data set typically generates numerous preliminary event hypotheses clustered in space and time around the location and origin time of the actual event. Extracting large events reduces the processing load and eliminates conflicts efficiently near the start of the processing sequence. This extraction is done immediately after the grid search (processes 2 and 3 in [Figure 17](#)).

A recursive procedure whose first step is a search for the event hypothesis with the most defining phases is used. If this event has more than a specified number of defining phases, the large event extraction procedure is attempted. The large event is first screened for inconsistencies in its phase associations. If inconsistencies are found, these events are referred to as degenerate hypotheses, and the proce-

cedure to remove the inconsistencies is the same as split analysis (see [“Analyzing Splits and Redundancy” on page 90](#) for the usual procedure to handle these events). A user-specified parameter controls the maximum number of splits from degenerate hypotheses that are processed. This parameter is an efficiency measure to limit the exhaustive exploration of all possible combinations for a degenerate hypothesis. When the estimated number of splits is less than this limit, all possibilities are examined. When the number exceeds the limit, the best hypothesis event is the only one that is located. The best hypothesis event is selected by the highest value for the pairwise chi-square association test at each station. The gain in efficiency obtained by bypassing the exhaustive expansion of degenerate events can be high because large events tend to generate many degenerate hypotheses.

The event, or events if several resulted from the split analysis, are then located (process 2 in [Figure 17](#)). If at least a specified number of defining arrivals (the same minimum number that is used to trigger the extraction) remain after location and outlier analysis, then the event is confirmed. Split and redundancy analysis is then performed on the remaining events (processes 3 and 4 in [Figure 17](#)). Phases that are associated with both the large event and another event are assigned to the large event.

ANALYZING SPLITS AND REDUNDANCY

Split and redundancy analysis identifies and corrects event hypotheses that have inconsistencies in the phase associations (process 4 in [Figure 17](#)). It also eliminates events with association sets that are a subset of another association set, under the restrictions outlined in the following paragraphs.

Eliminating Split Degenerate Hypotheses

The initial event construction can produce preliminary event hypotheses that include incompatible arrivals such as two or more arrivals at the same station identified as the same phase, or the same arrival identified as two or more different

phases. These cases are called “degenerate,” and when found, the event is split into two or more separate, self-consistent events. The event definition test is reapplied after the splitting because some of the resultant events may no longer pass the test (see [“Applying Event Definition Criteria” on page 86](#)).

Generating Alternate Regional Group Hypotheses

To allow multiple interpretations of phases within a regional group, several preliminary event hypotheses are generated when a regional pair is present. The generation of extra hypotheses follows these rules:

- Extra hypotheses are generated only when the regional group has two named phases.
- The following alternative phase interpretations are allowed: Lg may be interpreted as Sn or Rg; Sn may be interpreted as Lg; Rg may be interpreted as Lg.

Eliminating Redundant Event Hypotheses

The same set of associations can be consistent with two adjacent grid points. Similarly, a subset of associations can be consistent at an adjacent grid point. Consequently, a test is applied after splitting to remove these redundancies from the list of preliminary events. For an event to be deemed redundant, its arrivals must form a subset of another event's arrivals, and the common arrivals must have been identified as the same phase. In addition (unless an option to freeze arrivals at the beam points is selected) the *driver* must also be the same for both preliminary events. In the case of identical sets of associations, the one that best fits its grid cell is retained.

Complex Data Structure for Efficient Redundancy Check

A double linkage is established when arrivals are associated with events. Each event includes pointers to its set of associated arrivals, and, similarly, each arrival includes pointers to the events with which it is associated. The algorithm used to check redundancy takes advantage of this linkage, as illustrated in [Figure 23](#). Arrivals are numbered from 1 to 7 and events are labeled from A to E. The same arrivals are repeated at the top and bottom of the figure. Events include pointers to their associated arrivals, and arrivals include pointers to events with which they are associated. To test whether an event (referred to as *Event A*) is redundant with another, its associated arrivals are examined, and the arrival associated with the smallest number of events is identified. If that number is one, the event cannot be a subset of another. Otherwise, the other events connected to that arrival are examined and their arrival sets are compared to that of *Event A*. If the arrival set of *Event A* is found to be a subset of one of these events, then *Event A* is redundant. The double-linked structure supports a reduced search space. The link from arrival to events allows the search to be restricted to events linked through at least one arrival.

For efficiency, a partial redundancy analysis is performed during the association of corroborating arrivals if the number of formed preliminary events reaches a user-set limit. This redundancy analysis is only applied to preliminary events built since the prior redundancy check.

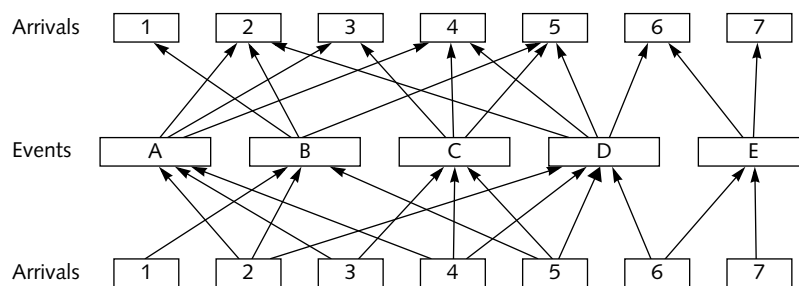


FIGURE 23. LINKS BETWEEN GA PRELIMINARY EVENT HYPOTHESES AND ARRIVALS

ANALYZING LOCATION AND OUTLIERS

After redundant hypotheses have been eliminated, all remaining event hypotheses are located and screened. Events are located by using the algorithms described in [“Event Location” on page 119](#). The locations are analyzed based on the residuals of travel time, azimuth, and slowness. Outliers are removed, redundancy checks performed, and the hypotheses are relocated (process 5 in [Figure 17](#)). Parameters control whether the location is computed at fixed or free depth. The fixed depth option uses the depth of the center of the grid cell. Even when free depths are allowed, locations are performed at fixed depth for events with a number of defining phases (*ndef*) less than a specified number. For events with an *ndef* larger than this value, when free depth location is requested, it is attempted first. If the location fails for free depth, a fixed depth location is performed.

Analyzing Outliers

A chi-square test is applied to the residuals of the time, slowness, and azimuth after location. Residuals for all data components are used simultaneously to compute a chi-square value. An outlier is defined to be an arrival whose chi-square value is larger than a specified threshold. The following procedure is applied to each preliminary event:

1. Eliminate the worst outlier, excluding the *driver*. First eliminate any instance of the worst arrival tagged `wc_restricted`; then eliminate any other defining arrival. Under any circumstances do not eliminate the arrivals tagged as `locked_association`.
2. If the event is now redundant, remove it from the list.
3. Relocate the event with the reduced set of arrivals.
4. Repeat the steps for the relocated event.

Redundancy Testing

A redundancy test is applied as a part of the location and outlier analysis. The test differs from the test applied after the association of preliminary events in that the *driver* is no longer required to be the same for both events.

Applying Event Confirmation Criteria

After the location is determined, a number of screening tests are applied to the event. These tests include the same weighted-count test that was applied after association of corroborating phases, a supplemental restriction that the event has a specified minimum number of associated arrivals, a restriction that the major axis of the error ellipse be less than a specified threshold, and an arrival quality test based on uncertainty of the slowness measurement (*delslo*) and on the distance between event and station. A probability-of-detection test may also be applied at this point (*max_obs_net_prob*) by using the computed event location and magnitude (see [“Estimating Magnitudes” on page 162](#)).

The network probability-of-detection test (which can be optionally applied to preliminary events both before and after the event location is computed) is intended to screen unlikely small events that were not detected at a sufficient number of probable stations. This test is only applied to events with a number of associated primary phases less than a specified limit.

The network probability-of-detection test determines whether or not the set of stations at which primary phases have been detected is compatible with the estimated location and magnitude of the preliminary event hypothesis. The probability of detection at each station is computed from the estimated location and magnitude of the event, the location of the station, nominal values of the station noise, the signal-to-noise detection threshold, and reliability (*nois*, *noissd*, *snthrsh*, and *rely* from the **siteaux** database table), and an amplitude attenuation table (*atten_file*). The *rely* values in the database may be overridden for all single stations and for all array stations. When this test is applied before location, the center of the grid cell is used as the estimated location, and the median of the individual

station magnitudes is used as the estimated magnitude. When applied after location, the computed location and magnitude are used. The stations in the network that are considered for the test are limited to stations that are active for the time period being processed.

The actual test takes the difference between the log of the product of the likelihoods for all stations in the network (equation 16) and its expected value (equation 17) and normalizes this difference by the square root of the variance of the expected value (equation 18). The preliminary event is eliminated if the expected value minus the actual value, divided by the square root of the variance, exceeds a specified threshold.

$$\sum_{k, \text{stations}} \begin{cases} \log P_k & \text{if a primary phase is detected at the } k\text{th station} \\ \log(1 - P_k) & \text{if a primary phase is not detected at the } k\text{th station} \end{cases} \quad (16)$$

where k runs over active stations and P_k is the probability of detection (see ["Detection Probability" on page 240](#)) at the k^{th} station times the estimated reliability of the station.

$$\sum_{k, \text{stations}} [P_k \log P_k + (1 - P_k) \log(1 - P_k)] \quad (17)$$

$$\sqrt{\sum_{k, \text{stations}} P_k(1 - P_k)(\log P_k - \log(1 - P_k))^2} \quad (18)$$

When applying the probability of detection test, only active stations are taken into account. Active stations have at least one detection in the considered time interval. Stations with arrivals that have been requested are also excluded from the test, because these stations would cause bias.

Hydroacoustic Blockage

The location subroutine does not consider blockage of hydroacoustic phases and may produce a solution with an error ellipse with no clear path to a hydroacoustic station with associated phases. When this is the case, all associations with completely blocked paths are removed from the event, and the event is relocated iter-

actively until either an acceptable solution is found, or the event is dissolved for lack of sufficient data. This is the last time that blockage is checked in the processing sequence before writing the solution to the database.

RESOLVING CONFLICTS

Conflicts, which are defined as one arrival tentatively associated with multiple event hypotheses, are resolved in numerous steps of network processing (processes 3, 6, 7, and 10 in [Figure 17](#)).

Two primary techniques are used in the conflict analysis: a cluster analysis and an association-based analysis.

Applying Cluster Analysis

The first method used in conflict resolution uses a clustering technique to identify and resolve conflicts between large events that share a high percentage of associated arrivals. This reduces a tendency to split large events during conflict resolution. Numerous small variations of a large event are frequently generated in GA. Cluster analysis provides a means of identifying a group of similar preliminary event hypotheses and reducing that group to the single “best event hypothesis.” The “selected hypothesis” is the one with the largest number of defining phases; the size of the error ellipse is used to break ties. The clustering method is only applied to clusters of large events (having many defining phases) so that all members have a high-confidence of representing the same event.

The key steps of the clustering algorithm are as follows:

- Select the preliminary event hypothesis with the largest number of defining phases that has not already been clustered. This number must be greater than or equal to a specified limit. If several event hypotheses have the same number of defining arrivals, then select the one with the smallest error ellipse. Call this event the “best event.”

- Form a cluster by identifying all other preliminary event hypotheses that have at least a specified percentage of associations in common with the “best event.” These associations must be time defining, meaning that the arrival time was used in the calculation of the event location. The phase identifications may be different for the different preliminary event hypotheses.
- Discard all preliminary event hypotheses in the cluster except for the “best event.”
- Continue as long as the preliminary event hypotheses have a sufficient number of defining phases.

Applying Association-based Conflict Resolution

Conflicts that remain after cluster analysis are resolved by an association-based method. Each arrival that is associated with more than one event hypothesis is assigned to the hypothesis that maximizes a weighted product of the goodness-of-fit and a measure of the quality of the event solution. The goodness-of-fit is based on time, azimuth, slowness, and log amplitude residuals. The event quality is based on the number of defining observations, size of the error ellipse, distance to the nearest station, and probability of detection, plus a factor that can be used to help retain small events. The test is applied iteratively, and all events are relocated, and all measures are recomputed after each disassociation of an arrival.

The conflicting arrival is assigned to the event hypothesis with the greatest quality measure:

$$L_{ij} = F_{ij}^a \cdot Q_i^{(1-a)} \quad (19)$$

where F_{ij} is a measure of the goodness-of-fit of the j^{th} arrival to the i^{th} event, Q_i is a measure of the quality of the i^{th} event, and a is a specified weighting factor. If a is zero then the goodness-of-fit of the arrival to the event solution will be ignored, and conflicts will be resolved in favor of the event with the highest quality. Con-

versely, if a is set to 1 then conflicts will be resolved only on the basis of goodness-of-fit. Both F_{ij} and Q_i are normalized between 0 and 1, so L_{ij} also varies in this range.

Goodness-of-fit

The goodness-of-fit of an association, F_{ij} , is based on the time, azimuth, slowness, and log amplitude residuals when available. A χ^2 value is computed as follows:

$$\chi^2 = \sum_j ((d_j - m_j) / \sigma_j)^2 \quad (20)$$

where d_j is the observed data, m_j is the theoretical data, and σ_j is the estimated standard deviation of the j^{th} datum. The χ^2 is unbiased by using the equation:

$$\chi^2 = \chi^2 \cdot N_{\text{tot}} / (N_{\text{tot}} - N_{\text{data}}) \quad (21)$$

where N_{tot} is the total number of data used to compute location and magnitude in the event hypothesis minus the number of model parameters in location and magnitude, and N_{data} are the number of data for the arrival in question. The quantity F_{ij} is the probability corresponding to this unbiased χ^2 value.

Event Quality

The quality of an event, Q_i , is the normalized weighted sum of two terms:

$$Q_i = (b \cdot Q_{1i} + c \cdot Q_{2i}) / (b + c) \quad (22)$$

The first term, Q_{1i} , is a measure of the quality of the event solution based on the number of defining phases, the size of the error ellipse, the distance to the nearest station, and the probability of detection. The second term, Q_{2i} , increases the event quality measure if the event is likely to be dissolved if the association is removed from it (if the number of defining phases is small). The term is introduced to allow a bias to preserve small events. The user-specified factors b and c allow adjustment of the relative weight given to these two terms.

The first term, Q_{1i} , can be considered a measure of how likely the event is to be real. It is computed as the following normalized weighted sum:

$$Q_{1i} = \sum_{k=1}^{NA} w_k \cdot M_k / \sum_{k=1}^{NA} w_k \quad (23)$$

where NA is the number of event quality attributes, w_k is the user-specified weight assigned to each attribute, and M_k is a membership function, which ranges between zero and one of how likely the event is to be real based only on the k th attribute. The M_k are approximated as linear ramps from 0.0 to 1.0 between interval bounds ($a1$ and $a2$) that are specified for each attribute ([Figure 24](#)). These values $a1$ and $a2$ are also called the no confidence boundary and the high confidence boundary.

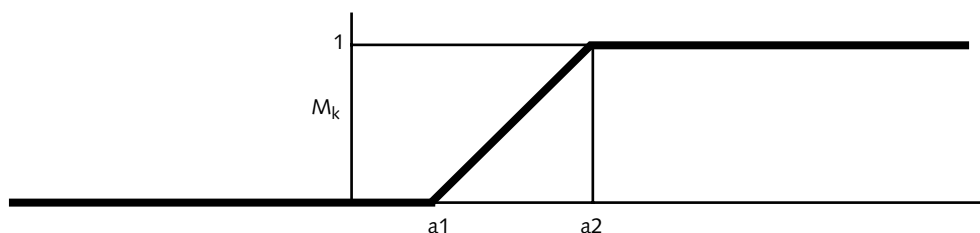


FIGURE 24. EXAMPLE OF MEMBERSHIP FUNCTION (M_k)

The event quality measure is computed differently for seismic or acoustic arrivals. The only event-based factors used for an acoustic arrival are the normalized number of arrivals of the same technology and the number of arrivals of technologies other than the arrival's technology. In this way the conflict resolution process is used to control whether a detection will be associated to the event with the most detections of the same technologies or to the event with the most detections of the other technologies.

The event quality member functions used to compute Q_{1i} for seismic arrivals, are as follows:

Network Processing ▼

- If an event hypothesis has a large number of defining phases then the event is more likely to be real than if it has a small number of defining phases. The interval bounds for M_k are specified for the no-confidence boundary (default is 3) and the high confidence bound (default is 10). The weight is also specified (default is 1.0).
- If the error ellipse is small then the network coverage is reasonable, and the event hypothesis is more likely to be real than if the error ellipse is large. The interval bounds for M_k are specified for the no-confidence boundary (default is 500 km) and the high-confidence bound (default is 10 km). The weight is also specified (default is 0.8, and the current configuration is specified as 0.05).
- If the nearest station is close, then the event hypothesis is more likely to be real than if it is far. The interval bounds for M_k are specified for the no-confidence boundary (default is 90 km) and the high-confidence bound (default is 10 km). The weight is also specified (default is 0.5, current configuration specifies as 0.0).
- If the network probability-of-detection estimate is consistent with the set of stations that detected the hypothesized event, then the event is likely to be real. The interval bounds are applied to the ratio of the residual used in the probability of detection event screening test and its standard deviation. Interval bounds are specified for the no-confidence boundary (default is 3.0) and the high-confidence bound (default is 1.0). The weight is also specified (default is 0.7, and current configuration is specified as 0.0).

As noted previously, the second term, Q_{i2} , is provided to control the probability of dissolving a small event and is defined in terms of the number of defining phases as follows:

$Q_{i2} = 0.0$ if the number of defining arrivals is greater than or equal to a specified value (default is 6).

$Q_{i2} = 1.0$ if the number of defining arrivals is less than or equal to a specified value (default is 3).

Q_{i2} is linear between these two bounds.

The event quality component is the sum of two membership functions each for hydroacoustic and infrasonic arrivals. The sum is computed as follows:

- To assign the arrival to the event that possesses the largest number of defining arrivals of the same technology, place the weight on the membership function that has interval bounds specified for the number of hydroacoustic/infrasonic arrivals in the event. The default bound values for the intervals are 0 and 4 for the number of hydroacoustic or infrasonic arrivals. The default weight is 1.0 for this membership function for both acoustic technologies.
- To assign the arrival to the event that possesses the larger number of defining arrivals of another technology (usually seismic), place the weight on the membership function that has interval bounds specified for the number of arrivals other than the technology of the arrival. The default bound values are 3 and 20 for this membership function. The default weight is 0.0 for this membership function for both acoustic technologies.

Conflict Resolution Procedure

The iterative procedure for applying the metrics described previously to resolve conflicts is initialized by computing Q_i for all events and L_{ij} for each of the conflicting associations. The following tasks are performed once for each event with conflicting associations:

1. Rank all conflicting associations based on their quality measure, L_{ij} . The rank is set to zero for the event with the highest L_{ij} , and it is greater than zero for all other events. Select the event with at least one conflicting association whose rank is zero and has the highest proportion of defining associations that it is likely to keep after conflict resolution. This proportion is $(n_1 + n_2)/n_{def}$, where n_1 is the number of defining associations that are not in conflict, n_2 is the number of conflicting

associations with rank equal zero, and $ndef$ is the total number of defining associations for the event. Resolve the ties by selecting the event with the highest event quality, Q_i .

2. For the current event, select the conflicting association with the highest quality measure, L_{ij} , and disassociate this arrival from all other events.
3. Relocate all events that have lost an association, and reapply the event definition criteria. Discard any events that no longer satisfy these criteria, and disassociate all of their arrivals. Recompute Q_i and L_{ij} for all events that have lost an association and re-rank all affected associations. Return to Task 2 if the current event has remaining conflicting associations with rank equal to zero.

Several iterations through this procedure may be required to resolve all conflicts. The following example provides such a case.

Example

Figures 25 and 26 illustrate a simple example of association-based conflict resolution. Four events have conflicting associations of three arrivals. Figure 25 shows the events, their associated arrivals, and the conflicts. Four events, A, B, C and D, have three associated arrivals in conflict, α , β and δ . Each square represents an arrival. The number next to the square is the rank of the association: 0 represents the best association for that arrival. The lines join the same arrival associated with different events. Assume for simplicity in this example that the ranking of associations is not modified after events are relocated.

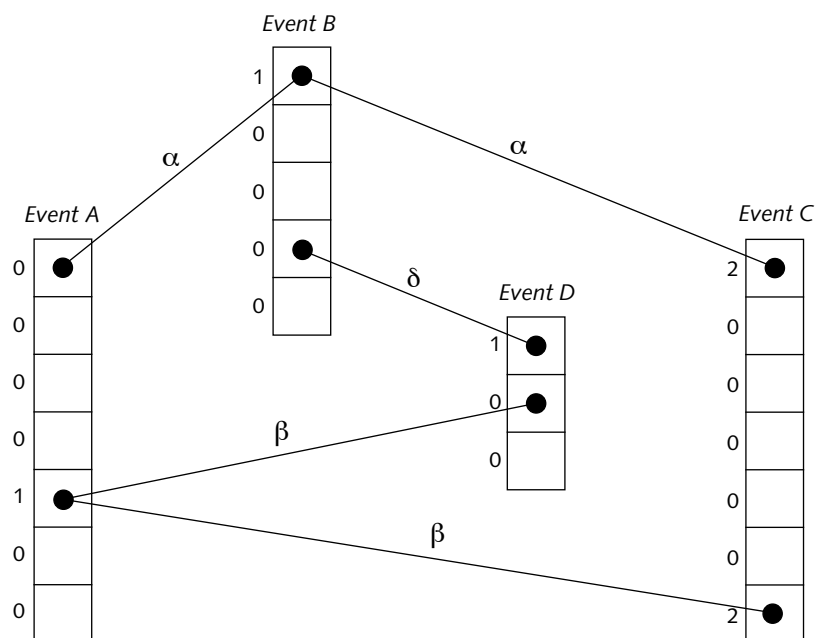


FIGURE 25. GA EVENT HYPOTHESES WITH ARRIVALS IN CONFLICT

Two iterations of the conflict resolution procedure are required to resolve all conflicts in this example. [Figure 26](#) illustrates the results at intermediate stages. The first iteration of tasks, which correspond to the conflict resolution procedure, are as follows:

1. Select Event A (Task [1](#) in the conflict resolution procedure).
2. Disassociate Arrival α from Events *B* and *C* (Task [2](#) in the conflict resolution procedure).
3. Relocate Events *B* and *C*; both events still pass event definition criteria ([Figure 26a](#)) (Task [3](#) in the conflict resolution procedure).
4. Select Event *B* (Task [1](#) in the conflict resolution procedure).
5. Disassociate Arrival δ from Event *D* (Task [2](#) in the conflict resolution procedure).

6. Abandon Event *D* because it does not satisfy event definition criteria (only one arrival left). Rerank conflicting associations of arrival β ([Figure 26b](#)) (Task [3](#) in the conflict resolution procedure).

Event *C* is not selected in the first iteration because it does not have any conflicting arrivals with rank equal to zero. Event *D* is not selected because it is dissolved during the processes of resolving the conflicts with Event *B*. The second iteration must resolve only one conflict:

1. Select Event *A* (Task [1](#) in the conflict resolution procedure).
2. Disassociate Arrival β from Event *C* (Task [2](#) in the conflict resolution procedure).
3. Relocate Event *C* ([Figure 26c](#)) (Task [3](#) in the conflict resolution procedure).

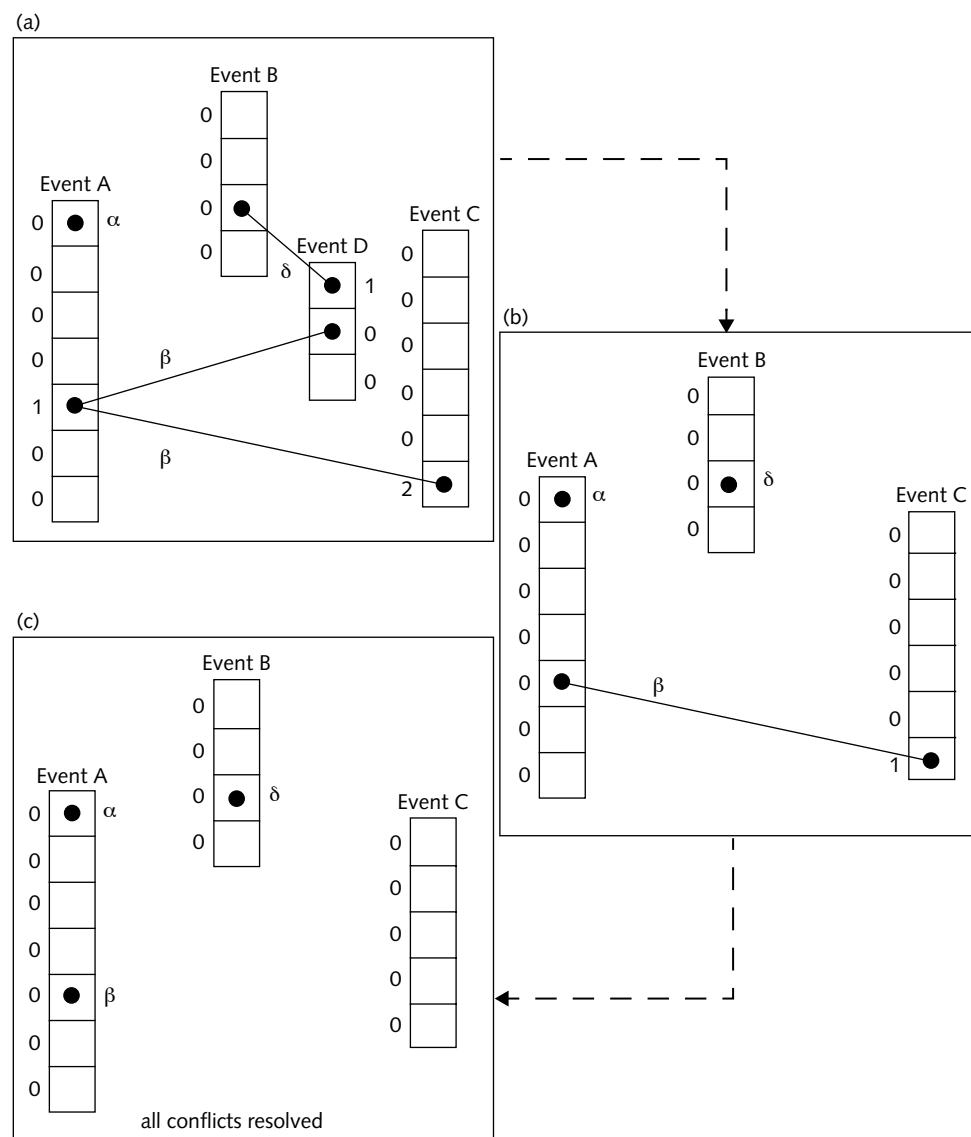


FIGURE 26. ASSOCIATION-BASED CONFLICT RESOLUTION PROCEDURE IN GA

All conflicts have been resolved after these two iterations, and Events A, B and C remain. The re-ranking after each conflict is resolved is an important step in the procedure. In this example, arrival β would have been disassociated from Event A if the ranks were not updated.

PREDICTING PHASES

Both defining and nondefining phases are predicted and compared to the associated phases for each event. If an event has a specific phase associated at a station, alternative associations of that phase to that event will be attempted.

Predicting Nondefining Phases

For each event and associated station, the travel time and predicted slowness and azimuth are computed for each phase in a specified list of predicted phases. A chi-square value is computed from the residuals between the observed and expected travel time, slowness, and azimuth for a predicted arrival. If the chi-square value falls within the specified limits, the predicted phase is added to the association set for the event (process 9 in [Figure 17](#)).

Several detections at a station can be identified as the same phase. Also, one detection can be identifiable as several phases. If either of these situations occur, the detection-phase pair with the best fit based on its chi-square value is selected and added to the association set.

Phases are not predicted outside of their theoretical distance ranges.

Predicting Defining Phases

The prediction of defining phases (process 8 in [Figure 17](#)) is similar to the prediction of nondefining phases. The main differences between prediction of defining and nondefining phases are as follows:

- The error ellipse is taken into account by commensurately enlarging the travel time, azimuth, and slowness windows for defining phases prediction, but it is not taken into account in the case of nondefining phases.

- Defining phases are predicted at stations that are not already associated, but nondefining phases are only predicted at associated stations.
- The event is relocated and outlier analysis performed if a new defining phase is added.

The initial state of the event must be preserved before location and outlier analysis, so that it can be recovered if the event is lost during relocation and outlier analysis. In this case, the newly added phases are removed and the initial state is restored.

Hydroacoustic Phase Blockage

When hydroacoustic phases are to be predicted, the blockage of the error ellipse of the event for which the phase is being predicted is first checked. No prediction is attempted if the ellipse is completely blocked.

CHECKING EVENT CONSISTENCY

Checks on the seismological logic of the event hypotheses are performed before writing the hypotheses to the database (process 11 in [Figure 17](#)). If a check requires modification of an event hypothesis, the event is relocated, and all of the tests are reapplied. Because of the complexity of the tests and the instability of the location of small events, this sequence can, in some circumstances, lead to an infinite loop. To avoid this situation, a maximum number of iterations through the tests is allowed. If an event exceeds the maximum number of iterations, it is written to the database even though it may not pass one of the required conditions.

Event consistency is checked as follows:

- *Deep event check.* Deep events are checked against the deep seismic cells to make sure that deep events are located in areas where deep seismicity is historically known to occur. One of the grid files contains the cell geometry involved in this check. If a deep event's confidence ellipse does not intersect any deep seismic cell, an attempt is made to relocate it at the surface. If this fails, the event is removed.

- *Distance-depth ranges.* Each seismic phase is constrained to fall within specified ranges in distance and depth. All associations are checked against these ranges. If an arrival is found to be out of its range, its phase name is changed to one that does fall in the correct range.
- *Regional pairs.* Phase combinations *Pn-Rg* and *Pg-Sn* are not allowed if no other regional phases are detected at the station. If such a combination is encountered at a station, the name of the secondary phase is changed to *Pn-Lg* or *Pg-Lg*.
- *Phases out of order.* This may happen when sufficiently large travel-time uncertainties for at least one of two phases exist. If two phases are found to be associated in reverse order, they are switched if they are both secondary. If the pair includes a secondary phase followed by a primary phase, the secondary phase is eliminated.
- *Local magnitude M_L outlier screening.* When a well-determined teleseismic bodywave network magnitude, m_b , is available for the event, the event is checked for large differences between the station M_L and the network m_b . The test is applied to events that have a network m_b uncertainty less than a specified value and where the m_b was estimated using a specified minimum number of stations. If the difference between the station M_L and the network m_b for a station is larger than either the maximum value or the specific maximum magnitude difference value for this station, the arrivals for the station are disassociated.
- *Nondefining phases residual screening.* The residuals of nondefining phases are screened using the same chi-square test as in the prediction module, using an outlier threshold.
- *Isolated secondary phases.* If specified, events are checked to ensure that no secondary phases are associated without a primary phase at the same station. If a secondary phase is associated without a primary phase, the secondary phase is disassociated.
- *Phase duplication.* Events are checked to ensure that the same phase is not duplicated at the same station. If a duplication is found, the phase with the best time residual is kept.

TAGGING ARRIVALS FOR SUBSEQUENT PROCESSING

The GA application can be configured in various ways to process prompt data, late arriving data, and analyst reviewed data. To manage the various mission-dependent configurations, a database table, `ga_tag`, serves as an interface between `GA_DBI`, the program that concentrates mission-specific requirements and the GA System, which will act on the data according to the tagging performed by `GA_DBI`.

Late Arriving Data

A pipeline can be configured with an efficiency parameter activated. This is not the current IDC configuration. This parameter will prevent events that have already been formed in a previous pipeline from being reprocessed. Only events that include arrivals that were not available in a previous pipeline will be formed in the current pipeline. Previous results and an indication of whether or whether or not an arrival has been completely processed at least once through a previous pipeline must be tracked. The arrival in the `ga_tag` table is tagged to track this information. An arrival is tagged as `aa_processed` when it has been completely processed at least once. The first time that an arrival passes through `GAassoc`, it is tagged as `assoc_first` if it lies in the time interval between `start_time` and `end_time`. This indicates that the arrival was available for processing as a corroborating phase at the time of the first instance of `GAassoc`. When the arrival passes through `GAassoc` again, and its time is between `start_time-origin_time_limit` and `end_time-origin_time_limit`, it is tagged as `aa_processed`, provided it had been tagged previously as `assoc_first`.

If the arrival was not present the first time `GAassoc` processed its time segment, for instance because it was late, it will not have been tagged as `assoc_first` and therefore will not be seen as completely processed in a subsequent network processing run.

The main purpose of tagging the arrivals as `aa_processed` is to be able to recognize previously processed events in later network processing runs and to avoid unnecessary reprocessing of these events. After the initial association loop in *GAassoc*, events that include only tagged arrivals are eliminated from the preliminary events list, because they have been fully processed in an earlier pipeline.

Auxiliary and Analyst Reviewed Arrivals

GA_DBI is a site-specific module designed to support mission-dependent requirements through tagging of arrivals. For instance, all arrivals from auxiliary stations at the IDC are tagged as `wc_restricted` and `probdet_restricted`. When an arrival is tagged as `wc_restricted` (weighted-count restricted), *GA* processes that follow from *GA_DBI* will not count the arrival in the event definition criterion (see [“Applying Event Definition Criteria” on page 86](#)). Similarly, an arrival tagged as `probdet_restricted` will not be used in the probability-of-detection test.

Hydroacoustic Arrivals

Arrivals identified as defining hydroacoustic phases (H phases) by station processing are tagged as `driver_restricted` under potential overload circumstances. The reason is that the absence of azimuthal information for these phases can cause a large number of preliminary events to be formed in the initial loop of event formation. Arrivals whose `snr` falls below a specified threshold will not be used as drivers. In addition, if the number of arrivals per hour available for use as drivers is higher than a specified number, the number of such arrivals is limited by taking the arrivals with the highest `snr`.

TIME WINDOWING AND TIME STEPPING

Processing occurs in each of the three automatic pipelines on limited time intervals and is triggered at regular time intervals. The current configuration is to process an interval of 40 minutes every 20 minutes, with an overlap of 20 minutes between successive intervals.

Hydroacoustic and infrasonic waves propagate more slowly than seismic waves. Hence the minimal length of the time window necessary to collect the data needed to form single-technology events for these two types of sensors is larger than the minimal length for primary seismic data. Because an objective of the IDC is to be able to form composite (multiple technology) events, multiple technology-dependent processing windows, which are adapted to a typical velocity for each physical phenomenon, must be supported.

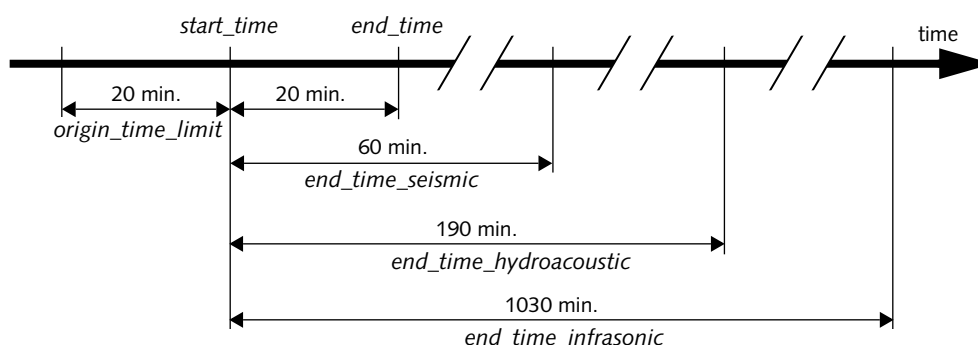


FIGURE 27. TIME-WINDOW PARAMETERS FOR GA DATA PROCESSING

[Figure 27](#) shows the organization of the time windows for the S/H/I technologies. The time windows determine the arrivals to be loaded as part of the exhaustive grid search (process 1 in [Figure 17](#)). The time interval between *start_time*–*origin_time_limit* and *end_time* is searched for seismic drivers. Corroborating seismic arrivals are sought for the seismic drivers in the interval between *start_time*–*origin_time_limit* and *start_time*+*end_time_seismic*. Hydroacoustic drivers and corroborating arrivals are sought in the interval between *start_time*–*origin_time_limit*

**Network ▼
Processing**

and *start_time+end_time_hydroacoustic*. Similarly, drivers and corroborating arrivals for infrasonic data are sought between *start_time–origin_time_limit* and *start_time+end_time_infrasonic*. Time increments for GA processing are controlled by the *DACS* system for all pipelines. The *end_time* of the previous interval becomes the *start_time* of the current interval.

For the purpose of resolving conflicts between different events, certain events that have been formed previously in the current pipeline or in a previous pipeline are loaded into the current processing interval (starting at process 7 in [Figure 17](#)). These events include arrivals in the appropriate current time window. Therefore, an event that is already in the bulletin can be modified during processing of a later interval. An event in the database can be considered stable when and only when the time of its latest currently associated arrival is less than *start_time–origin_time_limit*. This is strictly true in the case when no previous pipeline is used as input to the current pipeline.

When the output from a previous pipeline is used as input to the current pipeline, the interaction between the current pipeline and the previous pipeline must be considered. Events in the current pipeline that conflict with events from the previous pipeline having an arrival in the current processing interval are loaded into the current interval for processing. In some rare cases events whose latest associated arrival is less than *start_time–origin_time_limit* can be modified. An upper bound to the maximum time difference exists in that rare case as well, and it is a function of the maximum time between origin time and time of the latest possible arrival within an event. This maximum possible time is called *max_time*. The event will not be modified if the time of its latest associated arrival is less than *start_time–origin_time_limit–max_time*. If the maximum distance between infrasonic source and receiver is 60 degrees, *max_time* is about 6 hours.

A future version of the Network Processing applications will have the option to enforce strict limits on the origin times of events to be loaded into the current processing interval. These limits will eliminate the pathological cases described previously.

RETRIEVING AUXILIARY SEISMIC STATION DATA

The *Retrieve Subsystem* retrieves segmented raw waveform data from auxiliary seismic stations to refine estimates of seismic event hypocenters. Arrival time windows are predicted from initial hypocenters determined from the global set of primary seismic stations.

Due to the large number of stations involved, a search for detections in waveform segments from all auxiliary seismic stations for each event location is not cost effective. A set of rules is used to rank auxiliary seismic stations in terms of the likelihood that each station will detect the event signals and the significance of the data for improving location estimates. The program that evaluates the utility of each auxiliary seismic station for improving the location of an automated event is *WaveExpert*.

WaveExpert estimates the probability of detection at a particular station as a function of the predicted amplitude (which is determined differently depending on whether the station is at regional or teleseismic distances from the event), the station noise level, signal and noise amplitude variance, and a specified STA/LTA threshold for detection (see ["Detection Probability" on page 240](#)).

The use of auxiliary seismic station data for improving preliminary locations is estimated from an analysis of the covariance matrix of the combined primary/auxiliary seismic network. The analysis quantitatively measures the importance of the data from individual auxiliary seismic stations in improving epicenters and depths of solutions. The algorithm gives the marginal change in the estimated event location error from the inclusion of a auxiliary seismic station in the overall station set.

WaveExpert's main functions are to rank stations (allowing for user-specified ranking) and to request data intervals. These functions and methods for optimizing the *WaveExpert* parameters are described in the following sections.

Ranking Stations

A Jackknife algorithm is used to estimate how much an auxiliary seismic station interacts with the other auxiliary seismic stations that may be included in the final station set. In the following description of the algorithm, italicized words represent input parameters:

1. Form a core network, comprised of primary seismic stations associated with the event.
2. Sort auxiliary stations by probability of detection (see [“Detection Probability” on page 240](#)). Loop over the sorted auxiliary seismic stations having detection probability *prob_threshold* (currently 0.35).
3. Calculate the percent change in the volume of the error ellipsoid between the core network and (core network + current auxiliary station) assuming the current auxiliary station detects the event.
4. If percent volume change is above *err_vol_change_threshold* (currently 5%); add this station to the core network.
5. Go to the next station (repeat from Step 3).

The station rank value is set to the percent change in the volume of the error ellipsoid as determined previously. This value is compared with a user-supplied threshold to decide whether or not to request data from a auxiliary station.

WaveExpert also allows special user-defined rules that override the automatically determined selection. The user-defined rules are established from empirical analysis of auxiliary seismic station data based on seismic experience of station sensitivity for given source/receiver paths.

User-specified Ranking

If, for any reason, the Jackknife ranking method is unsatisfactory, user-supplied rules can be used to rank the stations. The rule-based rank values supersede values determined by the Jackknife algorithm. The Jackknife values are, however,

available for use in the rules, and if no rank is given by the user-supplied rules for a particular station, that station will maintain the rank value given by the Jackknife method.

The current user-defined rules allow data retrieval from up to 50 auxiliary seismic stations that are within 5 degrees of the event location.

Data requests will be issued at up to *max_stations* (currently eight) auxiliary stations. If user-defined rules are used, and the *add_rule_rankings* parameter is set, then the number of stations added by the user-defined rules will be added to *max_stations*.

After stations are ranked they are compared to a threshold. Any stations with rank values of at least the threshold value are candidates for auxiliary data request.

Requesting Intervals of Data

After the station ranking is complete, the request intervals for candidate stations are determined by using a list of specified phases and the lead and lag parameters *lead_time* and *lag_time*. Travel times for all phases are calculated by using specified travel time tables. The interval start time is as follows:

$(\text{event origin time}) + (\text{minimum phase travel time}) - \text{lead_time}$

Likewise, the interval end time is as follows:

$(\text{event origin time}) + (\text{maximum phase travel time}) + \text{lag_time}$

Only phases within the specific phase-distance limit will be considered for interval calculations. A phase distance limit is the distance range of a phase as given in its travel time table. If no specified phases are applicable, the "P" phase travel time is used by default.

Intervals for a station will be requested for all specified components (default list is sz, sn, se, bz, bn, be) that are also listed for the station in the **sitechan** table. If no **sitechan** entry exists, all components will be requested. Additionally, requests for an array station may be expanded to include each array element as listed for the array in the **affiliation** table.

To preclude duplicate station-interval requests, the **request** and **wfdisc** database tables are checked for both existing data and existing requested intervals, which coincide with the new request interval. Any existing data and/or request intervals are removed from the new interval before entry into the **request** table.

Optimizing Auxiliary Seismic Data Request Parameters

The objective in optimizing the auxiliary seismic data request parameters is to obtain a close match between the expected contribution of the auxiliary data as determined by the selection criteria and the actual value of the data that are obtained. Specifically, optimizing the auxiliary data requestor consists of maximizing the percentage of requests that yield defining data, while minimizing the percentage of detectable arrivals missed.

The probability of detection is based on the estimated amplitude, the station noise level, signal and noise amplitude variance, and the STA/LTA threshold for detection. Station noise level, signal amplitude variance, and noise amplitude variance (stored as *nois*, *noissd*, and *amcorsd*, respectively in the **siteaux** table of the database [IDC5.1.1Rev1]) can be estimated from noise and signal amplitude measurements. The STA/LTA threshold for detection is best obtained using the methods described in [“Optimizing Detection Parameters” on page 20](#).

The detection probability is used in conjunction with values of *data importance* to estimate the rank of the station, with respect to the other stations in the network. Data importance reflects the contribution of the station (in terms of arrival times, azimuths, or slownesses) to the event location. All the relevant auxiliary seismic stations are ranked for a given event, and requests are performed for those above a given ranking threshold up to a maximum number. Several parameters can be tuned to optimize this ranking process.

In addition to the default algorithm for station ranking, customized rules that use other attributes, besides detection probability and data importance, to determine the station rank can be applied to a given station. The attributes are azimuth, distance, and gap reduction. For example, if stations A1 and A2 have back-azimuths

of 10 and 90 degrees, and station B1 has a back-azimuth of 30 degrees, the gap reduction for B1 is 20 degrees. This added functionality allows azimuthal effects at the station to be considered. For example, the knowledge of azimuth ranges for which Lg blockage is observed could be incorporated.

Event Location

This chapter provides a description of the techniques by which the IDC locates terrestrial, oceanic, and atmospheric events.

- [Overview](#)
- [Inputting Data](#)
- [Establishing Initial \(Seed\) Locations](#)
- [Predicting Travel-times, Azimuths, and Slownesses](#)
- [Using Least-squares to Invert Data](#)
- [Evaluating Solutions](#)
- [Updating Event Locations](#)
- [Estimating Errors](#)

Event Location

OVERVIEW

The determination of event hypocentral locations is a core functional element of many automatic and interactive applications within the IDC system. Event locations are determined via an iterative non-linear least-squares inversion as originally introduced for seismic data by Jordan and Sverdrup [Jor81] and extended by Bratt and Bache [Bra88] to include azimuth and slowness data. One strength of this approach lies in its robust error estimation capabilities to exploit previous and current variance knowledge, either simultaneously or independently. A full regionalization capability for defining travel-time knowledge has also been integrated into this process.

The significant event location processing steps are shown in [Figure 28](#). The data to be inverted are first selected based on rules established for the use of phases (process 1). Unless an initial location is already available, an initial (seed) location is guessed as a starting point (process 2). The iteration cycle (processes 3 through 6) involves calculating the residuals between the measurements of travel time, azimuth and slowness and the values predicted from travel time models (process 3); inverting the data for a new location with or without damping (processes 4 and 4*) and then evaluating the solution for stability and convergence (process 5). During convergence the solution is updated (process 6) in each cycle of the iteration. If the solution is unstable or is not meeting the convergence criteria, the process is abandoned. If the solution is converging but has not yet met the convergence criteria, then a new iteration begins. When the convergence criteria have been met, the error ellipse is estimated (process 7).

The processing steps shown in [Figure 28](#) are described in the following sections. [Table 14](#) lists the software library that performs each of the event location processing steps. This library is called by several applications that are listed in the remaining sections of this overview.

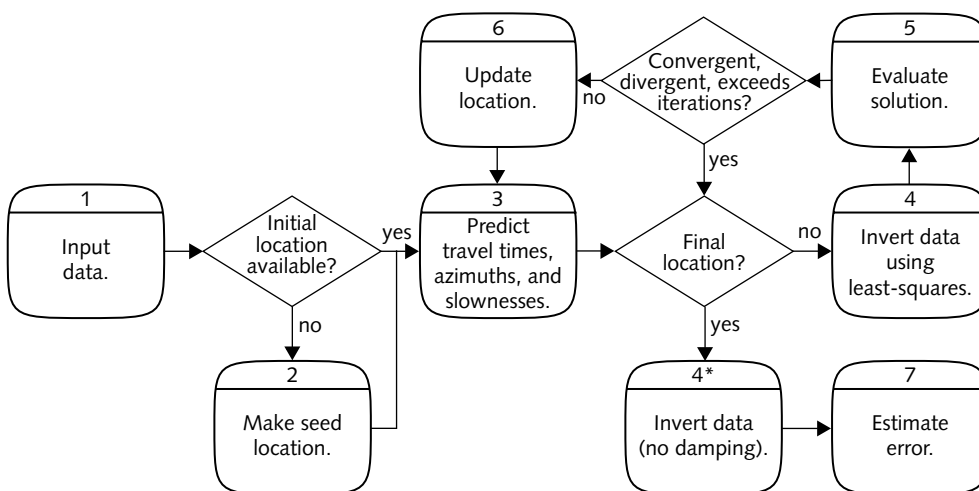


FIGURE 28. EVENT LOCATION PROCESSING

TABLE 14: SOFTWARE USED IN EVENT LOCATION PROCESSING

Process	Software	Reference(s)
1 Input data.	<i>libloc</i>	
2 Establish initial (seed) location.		
3 Predict travel times, azimuths, and slownesses.		
4 Invert data using least-squares (with or without damping).		
5 Evaluate solution.		
6 Update location.		
7 Estimate error.		

Applications

The event location processing described in this chapter is used by several IDC applications. The key applications include the automatic association *GA* application (see [\[LeB96a\]](#), and [\[LeB96b\]](#)), the *Analyst Review Station* program, *ARS* [\[Wan96\]](#), and the batch–event location application, *EvLoc* [\[Nag95\]](#).

Automatic Association

The event locator is a critical element in the process of resolving conflicting associations among the numerous event hypotheses originally produced by *GA* (see [“Resolving Conflicts” on page 96](#)). In this scenario, the starting (seed) location is the *GA* grid cell center in which the event was originally constructed. Such seed locations are generally good and convergence is almost assured after a good association set has been formed.

Analyst Review

When an analyst is reviewing the automatic association events with *ARS* for a given time segment, the seed location is the automatic association location. This location almost always provides the analyst with a good starting location. After event refinement by the analyst, the event is saved to the analyst-evaluated database account. Typically, the analyst will run the location program in triple-location mode (see [“Triple Location” on page 123](#)). Only the preferred location is output to the database account.

In some cases an analyst must build an event. Because no starting location exists, the seed (starting) location algorithm is used to obtain a reasonable initial epicentral location. The depth will always start at the earth’s surface in this case. These newly constructed events will be written to the same output database account as those initially created by the automatic association process.

Batch Event Location

EvLoc is an application that may be run in batch mode to compute a large number of event locations in a single instantiation. This application may be desired within a pipeline process, or more typically, for research purposes.

Triple Location

Both *ARS* and *EvLoc* can be executed in triple-location mode. During normal operational IDC processing, *ARS* is routinely run in triple-location mode. Only one hypocentral location per event association set may be retained. Therefore, only one event hypothesis can be brought into *ARS* for a given association set assigned to a single *evid*. Within the *ARS* application, however, up to three location hypotheses may be reviewed simultaneously. The three locations include an origin fixed at the earth's surface, a free-depth location, and an event location restrained as prescribed by the analyst (typically a fixed depth location other than the surface). When the analysts have completed their evaluation, they choose a preferred solution from these three possibilities. Only the preferred solution is written to the output database account.

INPUTTING DATA

As a first step ([Figure 28](#)), the data that can contribute to the location solution are identified. Phases used in the location process are known as *defining* phases. A defining phase may have one or more of three attributes that are useful in the location process: time, azimuth, and slowness. Time-defining phases are the most common. For seismic data azimuth and slowness defining phases are generally more useful from array stations than from 3-C stations because the standard deviation of the errors tends to be much smaller for array stations. All infrasonic stations of the IMS are arrays, so for infrasonic data azimuth and slowness information are available for all stations. None of the hydroacoustic data from the IMS are processed as arrays, so azimuth and slowness are not available for these data. All secondary phases are treated as independent data in the inversion, including seismic depth phases.

Constraints on the number and position of defining phases for all three technologies are used to ensure that the results are reasonable (see [\[IDC6.2.5\]](#)). These rules are not explicitly applied by the locator but rather by the analysts during analysis.

Inclusion of Azimuth and Slowness

Arrival time information is often the best information to use in event determination, but it is often unavailable in sufficient quantity. In such instances, azimuth and slowness information, as obtained from f-k or 3-C analysis, can be added to obtain more reliable event locations with smaller uncertainties. In particular, azimuth/slowness estimates are critical when a location is required at a single station. In their formal hypocentral inversion procedure, Bratt and Bache [\[Bra88\]](#), demonstrated the value of including azimuth and slowness estimates in the seismic event-location problem, especially with respect to regional locations with few or insufficient arrival time data. Azimuth is also important for locating events using infrasound data, where very precise azimuth measurements can be achieved and where the arrival time measures are of relatively minor value.

The inclusion of azimuth/slowness estimates tends to reduce the degree of nonlinearity in the linearized expansion. Because pre-determined uncertainty estimates associated with arrival time information are typically better than that of azimuth/slowness information, these latter data can actually prove counterproductive to the goal of accurate event location. The magnitude of arrival time errors is generally less than 1% of the overall travel-time path; azimuth/slowness estimates tend to be associated with uncertainties between 1% and 20% of their respective parametric spaces. In cases where abundant arrival time information is available, making the linear assumptions more valid, the importance assigned to azimuth/slowness estimates should be decreased or omitted altogether. In practice, the IDC analysts exclude azimuth and slowness estimates from seismic data when more than five defining arrival times are available to the hypocentral location process.

ESTABLISHING INITIAL (SEED) LOCATIONS

Due to the nonlinearity inherent in the iterative least-squares inverse location process, a reasonable seed (starting) location increases the likelihood of finding the global minimum desired. Seed locations are determined based on the quality of the available information at the time of processing. The locations are in three primary forms in preferential order: 1) a previously determined event location, 2) a location based on an exhaustive grid search as determined by the GA subsystem, or 3) an internal “best guess” algorithm built into the location library (process 2 in [Figure 28](#)).

In the rare situation when a best guess seed location must be made, the strategy is to use the combination of arrival time, azimuth, and slowness information possessing the greatest chance for an accurate starting location. The seed algorithm uses a prioritized list of location scenarios to find a starting location. The scenarios are matched, in order, to the available data until a successful starting location is found.

Each scenario in the following list requires a specific set of data, and some scenarios are particular to certain technologies. For instance, Scenarios 1, 2, and 4 require seismic data, and Scenarios involving azimuth cannot be applied to hydroacoustic data. Any of the scenarios can be used to find seed locations for seismic events. Scenarios 3 and 5 could be used to find seed locations for events that have only infrasonic data, and Scenario 5 could be used to find seed locations for events that have only hydroacoustic data. As more experience is gained, the seed algorithm will be modified to allow infrasonic-only and hydroacoustic-only events.

1. Use the time defining S–P time at the closest seismic station (yields a reliable distance) and the best-determined defining azimuth from a seismic station (the latter based on smallest historical standard deviation) to compute a seismic epicentral location.

2. Use various combinations of defining S–P times and defining P-wave arrival times from seismic stations to compute a seismic epicentral location. The three methods listed below usually yield good starting locations and are implemented in order:
 - Find the nearest (best) small circle crossing point given distances from the three earliest defining S–P times. The small circles each have a radius (measured as distance) determined from the S–P time, centered about the station. Choose a point half-way between the two nearest small circle crossing points as the seed location.
 - Use two defining S–P times to obtain two small circle crossing points (potential starting locations each with its own origin time). A third defining P-wave arrival time defines the smallest travel-time residual to one of the two crossing points. Choose the crossing point with the smallest residual as the initial epicentral location.
 - Use a single S–P time to estimate an origin time, and then find crossings points based on two nearest independent defining P-wave arrival times. Because an origin time can be estimated relative to the single S–P time, two additional small circles can be determined relative to the P-wave arrival times. Choose a point half-way between the two nearest small circle crossing points as the seed location.
3. Use the two best defining azimuths available from two different stations to define two great circles whose intersections are candidates for the seed location. Choose the location that is closest (on average) to the two stations. When azimuth differences are less than 10 degrees, the intersection is poorly determined and this method is not used.
4. Compute a seismic epicentral location by using the best P-wave azimuth (for direction) and slowness (for distance) from a single seismic station.
5. Put the epicentral location near (approximately 1 km) the closest station. Use azimuth information, if available.

PREDICTING TRAVEL-TIMES, AZIMUTHS, AND SLOWNESSES

Predicting travel-time, azimuth, and slowness (process 3 in [Figure 28](#)) is part of the iterative process of estimating an event location. During each iteration, the distances and azimuths between the stations and the current hypocenter location are calculated. Predicted travel times, azimuths, and slownesses, including their path-, source- and station-dependent corrections are determined along with the derivatives associated with these computations. The residuals for each type of measurement (travel-time, azimuth, or slowness) are computed as the difference between the measurements and the predicted values for those measurements.

Predicting Seismic Travel-time, Azimuth, and Slowness

Travel-time, slowness, and azimuth of seismic phases can all be used for locating seismic events. To predict the values that would be measured in the data from a particular station, these parameters are modelled. As the models improve in accuracy the modelling error is reduced, thereby reducing the size of the coverage ellipses for the events.

The seismic travel-time model is based on standard one-dimensional travel-time curves. Corrections may be applied to account for the earth's ellipticity, the material between the station and the model surface, gross differences in the regional geology that affect the arrival times for all phases, and source-specific effects due to lateral heterogeneity. As more knowledge is gained about the characteristics of the data from a specific station, the corrections previously described are successively implemented and improved.

The slowness model is based on the derivative of travel time with respect to distance. Regional corrections for source-specific effects due to lateral heterogeneity (if known) may be applied.

Azimuth modelling is based on the exact azimuth between the station and the hypothesized event location, corrected for source-specific effects due to lateral heterogeneity (if known).

The methodology for station- and source-dependent knowledge via separate correction terms provides an incremental approach whereby individual contributions can be measured against one another. Travel-time calculations are a combination of a local/regional (station-dependent) velocity model, an ellipticity correction, an elevation correction (with an independent sedimentary velocity defined), a bulk static station correction term, and a source-dependent station correction. By permitting multiple one-dimensional travel-time models to be specified for a given station/phase based on the local velocity structure, additional travel-time corrections are minimized. This improvement is particularly important for regional branches, because a local velocity model most likely represents the largest adjustment to the travel time. In many circumstances the inclusion of a local one-dimensional model and a single-valued bulk station correction term will be sufficient to describe the travel time. If no station-dependent velocity model, bulk station correction, or separate sedimentary velocity is specified, then the default IASPEI91 travel-time tables are employed [\[Ken91a\]](#).

Seismic Travel-time Tables

The IASPEI91 travel time curves are used as the global standard travel-time curves for seismic event location and travel-time prediction using short-period data [\[Ken91a\]](#). Standard one-dimensional travel-time curves are not universally applicable for accurate location and error estimation in a heterogeneous earth, however. Therefore, in addition to the standard IASPEI91 travel-time tables, regionalized station/phase-dependent travel-time tables may be employed in this system.

The travel-time tables contain travel-time parameterized by distance and depth. The modeling error associated with a given phase-dependent travel-time table is also explicitly specified [\[Nag96\]](#).

Bi-cubic spline interpolation is employed to extract the travel-time, slowness, and ancillary derivatives from the two-dimensional (distance/depth) travel-time tables ([\[Pre88\]](#)). This algorithm was chosen because of its dominant global character, which reduces the effect of local curvature as a given travel-time curve transitions from one branch to another. This effectively eliminates oscillatory solutions that

jump back and forth on one or more travel-time branches simply because their derivatives are discontinuous. Stability is gained without requiring some artificial form of regularization, like damping. Some localized value/derivative information is lost, but because branch crossover distances are generally poorly known anyway, this behavior is not wholly undesirable.

Long-period travel-time predictions for LR and LQ are made by tracing rays on a path-dependent basis across the earth's surface where group velocities can be specified on a latitude/longitude grid. These travel times are path dependent in that they are ray traced within each cell along a great circle surface path as specified by the individual cells group velocities. All such segments are added together to obtain the total path-corrected travel time. For details of the algorithm see Appendix D of [\[Nag96\]](#).

Ellipticity Correction

Ellipticity corrections to short-period travel-time estimates are applied via three coefficient terms, τ_0 , τ_1 , and τ_2 , stored within ellipticity correction tables at pre-specified distance/depth sample points [\[Dzi76\]](#). Given the co-latitude of the source as θ , and the distance and azimuth from the epicenter to the station as Δ and ζ , respectively, the ellipticity correction is given by equation (24):

$$\delta t = \frac{1}{4}(1 + 3 \cos 2\theta)\tau_0(\Delta) + \frac{\sqrt{3}}{2} \sin 2\theta \cos \zeta \tau_1(\Delta) + \frac{\sqrt{3}}{2} \sin^2 \theta \cos 2\zeta \tau_2(\Delta) \quad (24)$$

Ellipticity corrections are specifically computed relative to the IASPEI91 travel-time model but are used for all one-dimensional tables [\[Ken91a\]](#). Only a very large velocity model variation will affect these coefficients and the resultant ellipticity correction. Further, at near offsets where regionalized travel-time tables are used, the ellipticity corrections are very small. Additional details are included in Appendix C of [\[Nag96\]](#).

Elevation Correction

An elevation correction to the travel time of a seismic phase is computed by tracing a ray segment between the station (at some elevation above sea level) and sea-level based on a specified sedimentary velocity. If no sedimentary velocity has been specified, then a default sedimentary velocity of 5.8 km/s and 3.35 km/s is assumed for P-wave and S-wave branches, respectively.

Bulk Static Station Correction

A bulk static station correction to the travel time of a seismic phase is applied on a station/phase basis. If no specification exists, then no correction is applied. These corrections are typically teleseismic based and are meant to remove the bias in the background teleseismic travel-time model for a given station/phase pair. The assumption of linearity with a zero-mean Gaussian distribution strongly encourages the use of this correction for all primary teleseismic P-wave phases. Bulk static station corrections are less meaningful when relative to a regionalized one-dimensional travel-time table as the table is generally assumed to have a mean of zero.

Source-specific Station Corrections

Source-specific station corrections (SSSC) to the travel time of a seismic phase attempt to correct for the travel-time effects of lateral heterogeneity. SSSC tables are specified by a two-dimensional parameterization along latitudinal and longitudinal nodes ([Figure 29](#)). A background travel-time table is required for each phase with all corrections measured relative to this table. Source depth is not a definable element of this parameterization, so these corrections are applied equally for all source depths. Employing bi-cubic splines ensures that second derivatives are continuous everywhere. Therefore, defining a source region with a 0.0 equipotential boundary condition along its entire outer edge guarantees continuity across the boundary. As a general policy, all boundary nodes (indicated by open circles in [Figure 29](#)) should have 0.0 corrections. The gray area in the figure represents the tapered zone between the actual corrections and the boundary. This zone should

be made wide enough so that the spline does not overshoot, especially if the corrections are large along the edges. The station can be located either inside or outside of the source region. Because of the latitude/longitude parameterization, the specification of SSSC tables near the poles is discouraged.

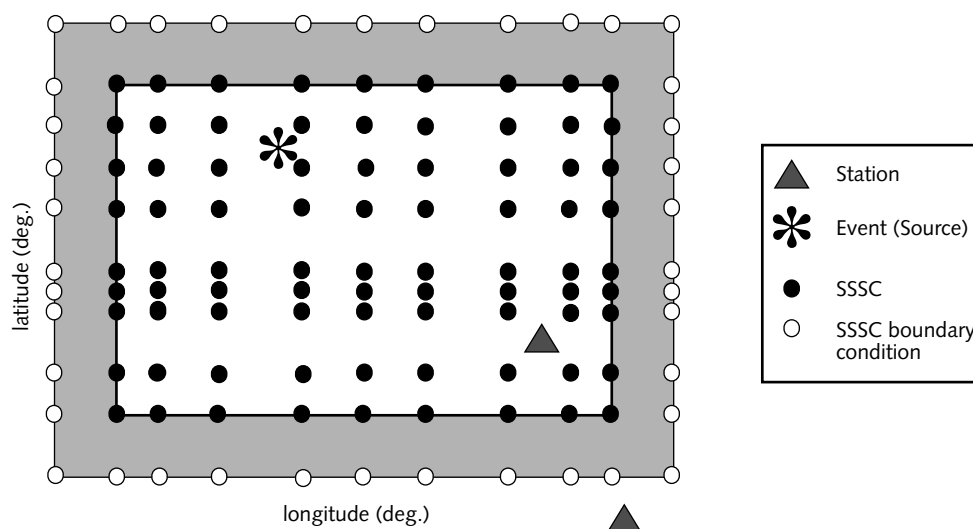


FIGURE 29. SSSC MODEL PARAMETERIZATION

By introducing SSSC information the phase-dependent modeling errors assumed within the one-dimensional travel-time tables should be reduced. A user may incorporate additional modeling error in one of two ways: 1) a single modeling error factor applied for the entire SSSC region, or 2) modeling errors sampled on the same latitude/longitude grid as the SSSCs themselves.

SSSC corrections should be based on data independent of the network, and initially, in lieu of other information, generic model-based corrections for a station may be attempted. Additional details regarding the SSSC structure can be found in Appendix E of [\[Nag96\]](#).

Slowness Prediction

Theoretical slowness (derivative of travel time with respect to distance) and slowness derivative determinations are computed using the bi-cubic spline interpolation scheme with distance/depth parameterized one-dimensional travel-time tables. The bi-cubic spline interpolation scheme requires computations of the second derivatives (in this case with respect to travel time). The first derivatives (slowness) are then easily derived from the sample values and second derivative information. Because of the nature of the group velocity curves for Pg, Lg, and Rg, slowness determinations are largely useless for these phases. The slowness for these phases should always be made nondefining.

Slowness/Azimuth Station Corrections (SASC)

Azimuth and slowness data from seismic stations are most useful for locating small events, especially when their data importances are high. Unfortunately, significant biases that can map into large event mislocations exist within these data. These biases, coupled with relatively large modeling and measurement errors, also produce very large error ellipses. The goal of the slowness-azimuth station corrections (SASC) is to remove these systematic biases from these azimuth-slowness measurements [\[Bon98\]](#). By simultaneously improving the previously determined modeling error estimates associated with these corrections, event locations with smaller confidence regions are achieved.

The slowness/azimuth station corrections and related modeling errors are represented on a binned polar slowness/azimuth grid (see Figure 26). Only those bins with sufficient available data from past events can contain a binned correction and modeling error. A constant correction and modeling error is defined over the entire bin. A background correction and modeling error can be defined for those bins without a specified SASC. The background correction is parameterized as a vector slowness correction (as opposed to a slowness/azimuth correction) to adjust for large-scale structural features near the station. An example of a station that requires a background slowness correction is TXAR where a clear systematic bias to the south is observed. [Figure 30](#) shows the slownesses and azimuths predicted

for TXAR based on network locations of events (triangles) along with the slownesses and azimuths observed in the TXAR data. The lines connecting the slowness/azimuth pairs are the vector slowness corrections required to correct the observations (from [Bon98](#)).

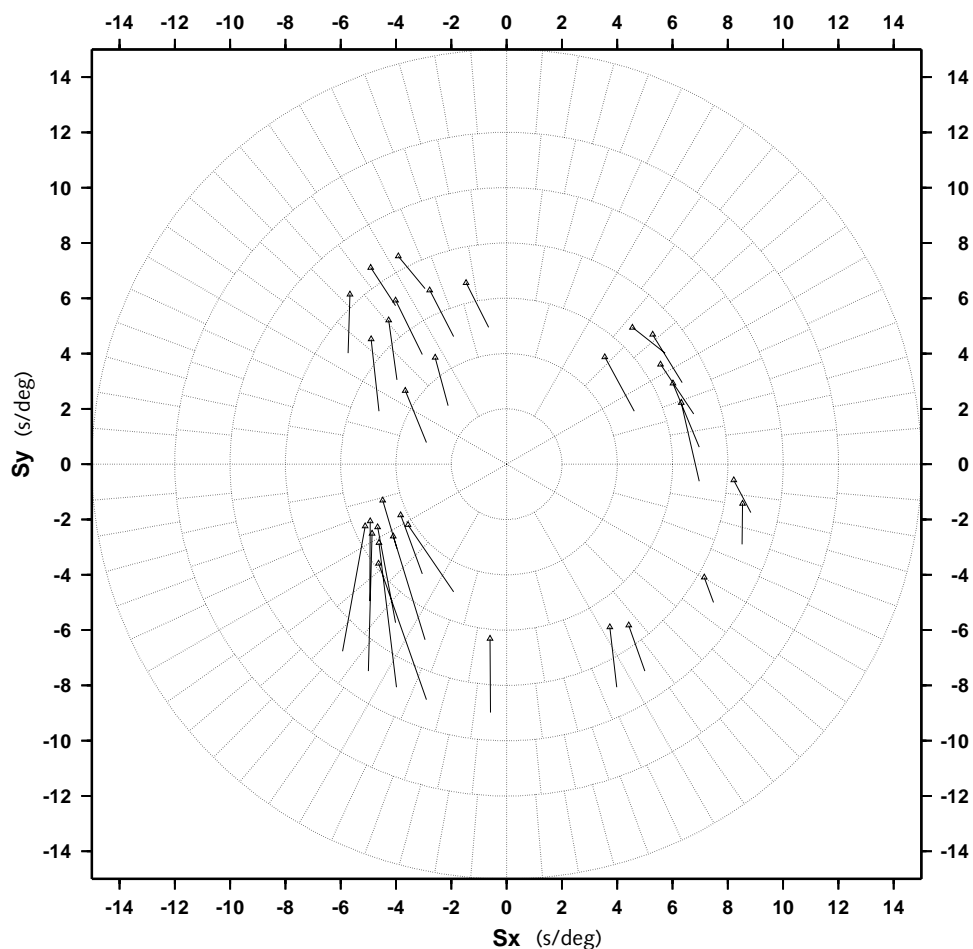


FIGURE 30. SLOWNESS/AZIMUTH STATION CORRECTION FOR TXAR

SASCs are corrections to observed data (namely, `arrival.delaz` and `arrival.delslo`) instead of model corrections like the SSSCs. Specifically,

Event ▼
Location

$$\begin{aligned} \text{slowness}_{\text{corrected}} &= \text{slowness}_{\text{observed}} - \text{slowness}_{\text{correction}} \\ \text{azimuth}_{\text{corrected}} &= \text{azimuth}_{\text{observed}} - \text{azimuth}_{\text{correction}} \end{aligned} \quad (25)$$

As with the arrival time errors, the measurement and modeling errors are separated in the slowness domain. The measurement errors are estimated during *f-k* or polarity analysis, while the modeling errors are specified in the SASC tables. Each bin with a correction must have a modeling error that is defined as the standard deviation (l1-norm) of the slowness and azimuth residuals.

The total error of slowness measurements for arrays is currently calculated as follows:

$$\text{del slo}^2 = \text{fkds}^2 + \frac{((\text{fkdk})/(\text{frequency}))^2}{\text{fstat}} \quad (26)$$

and for 3-C stations it's calculated as follows:

$$\text{del slo}^2 = \text{polards}^2 + 0.5 \cdot \text{polardk}^2 \cdot (1 - \text{rectilinearity}) \quad (27)$$

where *fkds* and *polards* represent the estimated modeling errors, and the rest of equations describe the measurement errors. With the introduction of SASCs, the modeling error attributes, *fkds* and *polards*, are specified within the SASC tables. The total error of azimuth estimates will continue to be calculated as follows:

$$\text{del az} = 2 \cdot \sin \frac{\text{del slo}}{2 \cdot \text{slo w}} \cdot 180 / \pi \quad (28)$$

The slowness and azimuth observations along with their associated residuals must have a consistent representation within the database as do the time domain corrections. Therefore, the uncorrected slowness and azimuth measurements are stored in the *slow* and *azimuth* fields of the **arrival** table, so that the slowness and azimuth residuals in the **assoc** table are relative to the corrected values as follows:

$$\begin{aligned} \text{slowres} &= \text{slowness}_{\text{observed}} - \text{slowness}_{\text{predicted}} - \text{slowness}_{\text{correction}} \\ \text{azres} &= \text{azimuth}_{\text{observed}} - \text{azimuth}_{\text{predicted}} - \text{azimuth}_{\text{correction}} \end{aligned} \quad (29)$$

Predicting Hydroacoustic Measurements

The hydroacoustic stations of the IMS are not arrays, so azimuth and slowness information is not obtained from their data. Travel-times of hydroacoustic phases and knowledge of all possible source locations for those phases (expressed as blockage grids for each hydroacoustic station) are used for locating hydroacoustic events.

The hydroacoustic travel-time model is complicated by the dynamic nature of the oceans through which the hydroacoustic phases propagate. Unlike the seismic travel-time predictions for which corrections are applied to a standard one-dimensional model, seasonally varying travel-time models are used to predict the propagation times for hydroacoustic phases. One-dimensional models are used only if the seasonal models are unavailable.

Hydroacoustic Travel-time Tables

Constant velocity travel-time tables may be used to approximate propagation times in the ocean. The default travel-time tables use a velocity of 1.485 km/s. However, because of the slow propagation velocity in the ocean and spatial and seasonal variations in temperature that induce velocity variations, actual travel times may depart substantially from constant-velocity travel times. To account for these spatio-temporal variations, the system provides for the use of seasonally varying, azimuth, and range-dependent travel-time tables specifically calculated for each hydroacoustic station. The travel times and their modeling uncertainty are specified on a polar grid centered at the station.

The travel times included in the hydroacoustic travel time tables are obtained from mode propagation calculations performed by program *ASPM* [Bur94] using a U.S. Navy water velocity database with a resolution of 5 nautical miles. The calculations are conducted for a point source at the actual depth of the hydrophone and a receiver close to the surface. The value of the travel time specified in the tables is based on the probability-weighted time algorithm (see [“Probability Weighted Time” on page 227](#)) for the multi-mode calculation and is obtained assuming a

Event ▼
Location

background uncertainty of 15 db re 1 μ Pa. Travel time differences between a two-dimensional seasonal model and a constant (1.485 km/s) model vary by tens of seconds and may just exceed 100 seconds for the very longest paths. One set of travel time tables may be specified per season, month, or any appropriate time division within a year. The default is to provide seasonal tables. Time boundaries for each season are adjustable.

The same travel time tables are used for the two types of phases, H and T, identified by hydroacoustic station processing (see [“Detecting Signals” on page 11](#)). Because the location of the coupling interface between underground and underwater propagation for T phases is ill-determined, the travel time for T phases is more difficult to determine with the same degree of accuracy than for H phases. This uncertainty is accounted for by adding a modeling error term to T phases to the modeling error values read from the two-dimensional tables.

Hydroacoustic Blockage

Hydroacoustic phases propagate long distances with limited attenuation in the water column. The phases are converted at the water-land interface of sizable islands or continents, but the converted waves do not propagate efficiently and are effectively blocked by the island or continental mass. Blockage is modelled for each hydroacoustic station as the maximum distance of propagation along a set of azimuths. A graphical illustration of the effect of blockage on a station is given in [Figure 31](#). For each azimuth a line extends from the station to the maximum distance for which the path is clear.

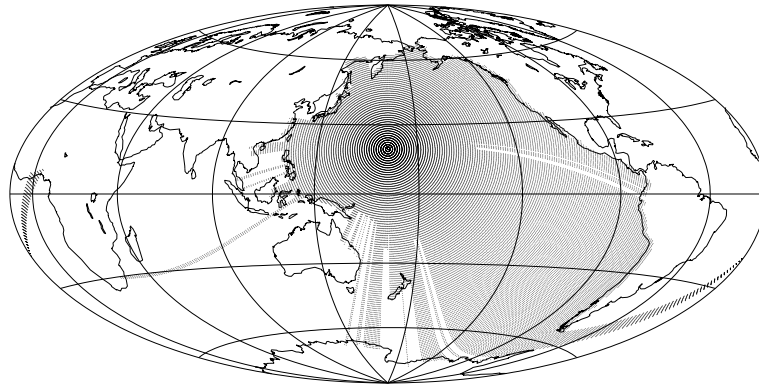


FIGURE 31. BLOCKAGE FILE FOR STATION WK30

The blockage subsystem is not used as an integral part of the location algorithm. Rather, it is invoked directly by several applications (*ARS* and *GA* for instance) to verify the status of a particular path. The event's error ellipse is considered when deciding whether or not a path from an event to a station is blocked.

Predicting Infrasonic Measurements

All of the infrasonic stations of the IMS are arrays and thus the data from these stations yield travel-times, azimuths, and slownesses.

Event ▼
Location

Infrasonic Travel Times

A single phase name, I, is currently used to model infrasonic propagation. The I phase propagation is modeled using a constant velocity of 0.320 km/s with a large modeling error. Similar to the constant velocity hydroacoustic travel-time tables, the travel time is specified as an extension of the IASPEI91 travel time curves. As more operational experience is gained, travel time differentiation between different phases, such as direct tropospheric (Iw), stratospheric (Is), and thermospheric (It) phases, will be attempted.

Infrasound propagation is a dynamic phenomenon depending upon the atmospheric conditions at the time of propagation. The travel time is dependent upon the dominant wind direction along the path sampled by the trajectories, which introduces potentially severe anisotropy in the travel-time maps. A planned upgrade to the infrasonic travel-time system will account for the dominant wind directions, which will be modeled monthly.

Slowness/Azimuth Station Corrections (SASC)

Infrasonic slowness and azimuth corrections may be made in the same manner that these corrections are made to seismic data (see ["Slowness/Azimuth Station Corrections \(SASC\)" on page 132](#)).

USING LEAST-SQUARES TO INVERT DATA

Hypocentral inversion or event location is performed as a nonlinear least-squares inversion of travel time, azimuth, and/or velocity data. An event location, confidence bounds, residuals and data importances [Min74] are computed using arrival time, azimuth and velocity (slowness) measurements from stations at local, regional and teleseismic distances (process 4 in [Figure 28](#)).

A sensitivity matrix, A , and a data residual vector, r , are required for the inversion. A is a partial derivative matrix of N data by M parameters that links the data residuals to an adjustment (correction) solution vector ($\Delta \mathbf{x}_e$) as,

$$\mathbf{r} = \mathbf{A} \Delta \mathbf{x}_e \quad (30)$$

Before solving for $\Delta \mathbf{x}_e$ the derivative matrix, A , and the residual vector, r , are row normalized (weighted) by their pre-determined errors (an inverse rms combination of measurement and modeling errors) to yield a weighted system matrix, A_w , and weighted residual vector, r_w . A_w is also column normalized (every column is made to sum, in an rms sense, to 1.0). This normalization numerically conditions the matrix so that each model parameter is treated equally. Column are “unweighted” after the inversion is complete so that the solution vector has its proper physical units.

With the sensitivity matrix and residual vector properly conditioned, the least-squares inversion can be performed. The Singular Value Decomposition (SVD) technique [Gou89] is applied to obtain a solution vector of location adjustments, $\Delta \mathbf{x}_e$:

$$\Delta \mathbf{x}_e = (\mathbf{A}_w^T \mathbf{A}_w)^{-1} \mathbf{A}_w^T \mathbf{r}_w \quad (31)$$

The singular values and vectors needed to construct the desired model covariance and data resolution matrices, C_m and R_x , respectively are also computed. Until convergence is achieved these two matrices need not be constructed (see [“Estimating Errors” on page 143](#)). If the condition number (the ratio of the largest and smallest singular values) is too great, then the location process is terminated. This situation is seldom encountered unless the solution is extremely poorly conditioned. In such cases, termination is justified and often highlights that the association set is inappropriate.

Damping can be included in the inversion procedure by adding a percentage of the largest singular value to all elements of the singular value vector. Fairly robust damping levels result from a choice based on the condition number that results from the inversion. The condition number is the ratio of the largest singular value divided by the smallest singular value. Fewer than 5% of all REB event locations

Event ▼
Location

require any damping whatsoever. Those solutions that do have damping have 1%, 5%, or 10% of their largest singular value applied. Solutions requiring more than 10% damping usually have serious problems that often result from a bad association set.

EVALUATING SOLUTIONS

Events are located through an iterative procedure (processes 3 through 6 in [Figure 28](#)). The procedure is similar to that used by other hypocentral location programs and is continued, while stable, until convergence is achieved, divergence is recognized, or the maximum number of iterations has been exceeded (process 5). Checking stability includes the tests for stability, convergence, divergence, and number of iterations. The algorithm to check the stability of the solution has been designed to exploit past empirical and intuitive knowledge in an attempt to salvage or stabilize the iterative process when potential problems are identified.

Stability

If the location adjustment at a given iteration is extremely large (namely, > 3000 km during the first few iterations and > 1500 km thereafter), then the revised location may be suspect; the stability of the solution is questionable. If the adjustment has been identified as being very large, then the first step is to scale down the adjustment. In the early iterations (the first 20% of the maximum iteration setting) this means not permitting the location to move more than 3000 km in a single step. For latter iterations this step is reduced to 1500 km. If large changes are required, this step slows down convergence but does not jeopardized the end results. Sometimes a rare numerical instability is introduced during a single iteration and causes the location to move dramatically away from any minimum. In this case, convergence itself is at jeopardy. Usually, down-scaled adjustment is enough to remedy the problem.

Divergence Tests

Two types of divergence are recognized. Local divergence is associated with a single iteration where the location seems to be getting worse because the current adjustment vector is larger than that encountered during the previous iteration. Global divergence is established when local divergence of sufficient magnitude is identified over two consecutive iterations. A single local divergent iteration is harmless and permits the iterative process to continue. A solution may commonly encounter a single iteration where conditions seem to be getting worse. But empirical evidence supports the view that when two consecutive locally divergent steps are identified, the last more than 10% worse than that two iterations prior, then failure is eminent. In this case, global divergence is declared, and the iterative location process is terminated with an appropriate error message.

Convergence Tests

Provided the solution is not globally divergent, assessment of convergence is evaluated next. Two types of convergence criteria are defined: 1) primary convergence criteria, and 2) a single auxiliary convergence criterion. If the higher-quality primary convergence criteria cannot be satisfied, then the single lower-quality auxiliary convergence test is used, but only after at least 75% of the maximum number of iterations permitted have been completed. Convergence cannot be attained until at least four iterations have been completed.

The most important convergence measure employed is based on the mapping details between the sensitivity (derivative) matrix and the data residual vector. This test uses a quantity, CNVGTST, based on the allowable perturbations in the data. Convergence is evaluated at the end of every iteration and is ideally constrained by the accuracy of the input data. [\[Pai82\]](#) characterize this test as follows:

$$\text{CNVGTST} = \frac{\|A_w^T r_w\|}{\|A_w\|_F \|r_w\|} < \text{ATOL} \quad (32)$$

where, ATOL, is a predetermined (empirically based) tolerance. CNVGTST is directly related to the input data accuracies, in this case relative to a unit-column normalized and weighted \mathbf{A} matrix. $\|\mathbf{A}_w\|_F$ is the Frobenius norm of \mathbf{A}_w . As a primary convergence measure, $\text{ATOL} = 1.0\text{e}^{-8}$, and as an auxiliary convergence measure, $\text{ATOL} = (1.0\text{e}^{-8})^{1/2} = 1.0\text{e}^{-4}$. The numerator in equation (32) yields a derivative vector. By taking its norm, an idea of the overall size (magnitude) of the derivative, dependent on both \mathbf{A} and the data, is obtained. Provided that the \mathbf{A} operator is unit-column normalized, the denominator maps the distribution of residuals to the overall system of equations. The above condition is therefore based on the allowable perturbations in the data coupled with the system equation. This convergence test has proven to be an extremely robust.

The remaining primary convergence tests are straightforward. If the total spatial adjustment vector in all dimensions is less than 0.5 km coupled with an origin time change of less than 0.5 s, then the solution has converged. Also, if the total weighted rms residual is less than 0.001, then the solution has converged. Only one auxiliary convergence test is available and it is only used when over 75% of the maximum number of iterations have already been exhausted. In this case CNVGTST is relaxed such that ATOL is reduced to 1.0e^{-4} . If ATOL is reduced for all solutions, premature convergence may be encountered in the early iterative steps.

If neither convergence nor divergence has been identified by the time the maximum number of iterations is encountered, then the location process finishes normally assuming normal convergence has been achieved; however, a warning code is returned to the calling application. The calling application determines what to do with these solutions. For example, the GA subsystem retains these solutions for further analysis, and ARS reverts back to the previously-determined location, if one exists.

Maximum Iteration Test

If convergence has not been achieved, iteration is terminated when the maximum number of iterations is reached. The maximum number of iterations is set to 60 for analyst review. A minimum of four iterations are always performed.

The iteration number is also used to control the conditions of the inversion. For seismic events the depth element of hypocentral parameters is constrained (fixed) during the first two iterations. If the depth of a seismic event is above the earth's surface, then it is labelled an *airquake* and is reset to zero. Four airquake results during the iterative process cause the depth to be permanently constrained to the surface for all subsequent iterations. Similarly, free-depth seismic event solutions that migrate to depths greater than 700 km are constrained to 700 km after four such occurrences (approximating the maximum depth of the deepest credible earthquake at 700 km).

UPDATING EVENT LOCATIONS

After all of the stability and convergence tests have been completed the event location is updated (process 6 in [Figure 28](#)) based on the newly-determined adjustment solution vector; $\mathbf{x}_{\text{new}} = \mathbf{x}_e + \Delta\mathbf{x}_e$. When convergence or divergence has been declared, the prediction process has to be invoked one last time so the data residuals and network statistical measures coincide with the most recently adjusted event location. In this case, $\mathbf{x}_{\text{final}} = \mathbf{x}_{\text{new}}$.

ESTIMATING ERRORS

Estimating errors (process 7 in [Figure 28](#)) includes calculating the model covariance matrix, \mathbf{C}_m , data resolution matrix, \mathbf{R}_x , the resultant 90th percentile error ellipse, depth error, and origin time error. The singular values and vectors necessary to build \mathbf{C}_m and \mathbf{R}_x are obtained from the inversion of the data.

The model covariance matrix, \mathbf{C}_m , can be directly extracted from the inverse decomposition described by equation (31):

$$\mathbf{C}_m = (\mathbf{A}_W^T \mathbf{A}_W)^{-1} \quad (33)$$

The covariance matrix yields the two-dimensional 90th percentile error ellipse and one-dimensional errors for depth and origin time. Assuming convergence has been reached in the location process, an error ellipse geometry can be prescribed

Event Location ▼

as points about the final location. At a given p percentile confidence level, each ellipse point, $\mathbf{x}_{\text{ellip}}$, is defined on an ellipsoid about the event location, $\mathbf{x}_{\text{final}}$, as follows:

$$(\mathbf{x}_{\text{ellip}} - \mathbf{x}_{\text{final}})^T \mathbf{C}_m (\mathbf{x}_{\text{ellip}} - \mathbf{x}_{\text{final}}) = \kappa_p^2 \quad (34)$$

where the confidence coefficient, κ_p^2 , assuming exact prior knowledge is a p percentile χ^2 distribution with M degrees of freedom, $\chi_p^2[M]$. An error ellipse computed assuming exact prior knowledge is a “coverage ellipse.”

The diagonal elements of \mathbf{R}_x provide the data importances, which indicate how important a given datum is to the overall solution.

Phase-dependent modeling errors, if addressed properly, improve the accuracy of the error-ellipse estimates and allow separation of modeling error from measurement error. These modeling errors may be specified by a single value, in a distance-dependent manner, or a distance/depth-dependent fashion as described in [“Predicting Travel-times, Azimuths, and Slownesses” on page 127](#).

Additional details on the computation of error ellipses where exact prior knowledge is not assumed can be found in [“Event Location Error Estimation” on page 252](#).

Analyst Review

This chapter describes how the IDC analysts review and improve the results of automated processing and includes the following topics:

- [Overview](#)
- [Reviewing Automated Processing Results](#)
- [Defining Events](#)
- [Techniques](#)

Analyst Review

OVERVIEW

The purpose of interactive analysis is to review and correct the results of the final stage of automatic network processing embodied in Standard Event List 3 (SEL3). This task involves retiming, renaming, and measuring the characteristics of phases identified by the automated processing, associating additional phases to the automated results, and scanning for events that were missed by the automated system. Analysts use the ARS application to perform these tasks [\[IDC7.2.2\]](#).

[Figure 32](#) shows the steps involved in analyst review. The period of data that is to be reviewed (usually a 24-hour period) is divided into segments of a few hours. Analysts select a time period for analysis (process 2 of [Figure 32](#)), select an event to review (process 3), review the automated results according to the type of event, correcting and rejecting events as appropriate (processes 4–6), check the quality of the event (process 7), and search for aftershocks if the event was large (process 8). New events are selected and reviewed (processes 2–5) until all of the automated results have been reviewed for the time period. At that point the quality of all of the events in the time block are reviewed (process 9), a scan is made of the unassociated phases in an effort to find events that were missed by the automated processes (process 10), and the quality of the events added during the scan is checked (process 11). A new time period is then selected for analysis (process 2), and interactive review continues until all of the time segments have been reviewed. At that point, the entire 24-hour period is checked for quality (process 12). The results of analyst review are published in the REB.

These steps are shown in [Figure 32](#) and are described in the sections of this chapter. Details of all of the steps are provided in [\[IDC6.2.5\]](#).

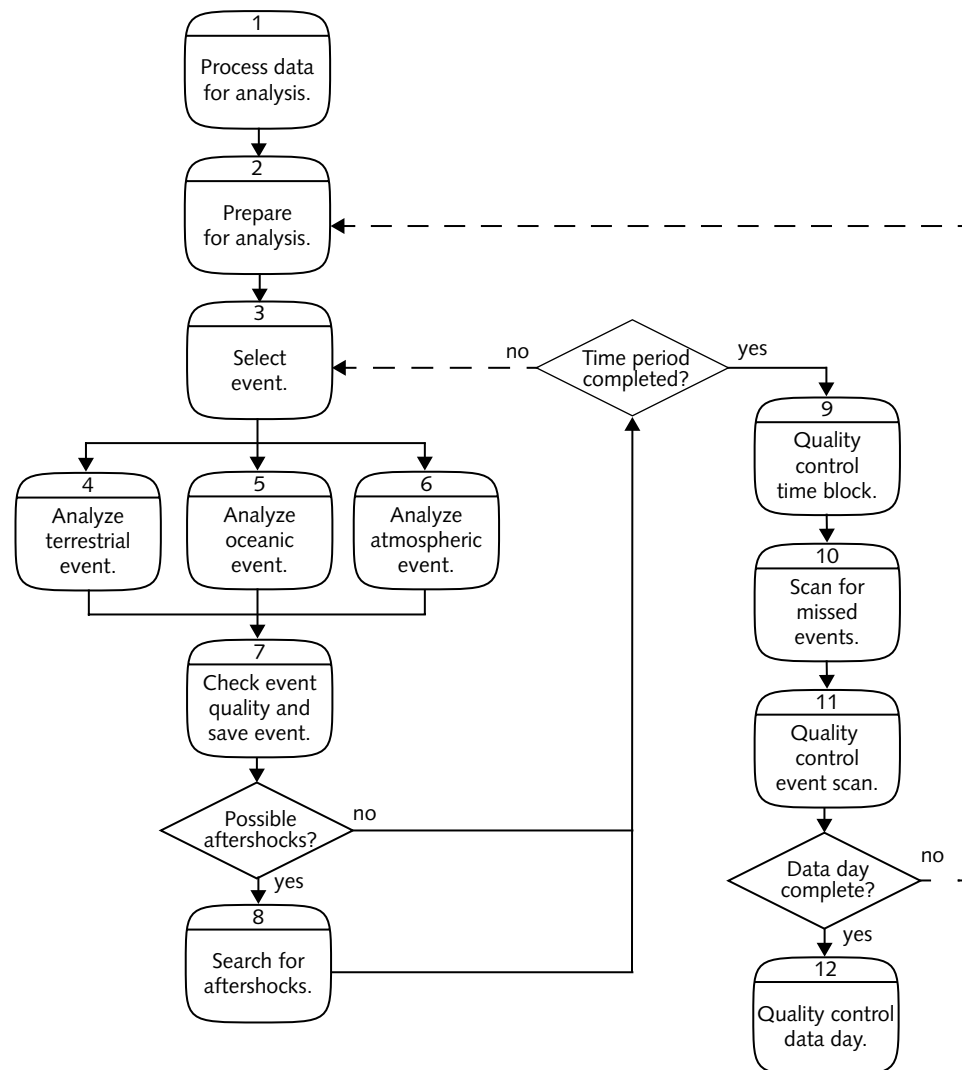


FIGURE 32. S/H/I INTERACTIVE ANALYSIS PROCESS

REVIEWING AUTOMATED PROCESSING RESULTS

Reviewing automated processing results (processes 2–6 in [Figure 32](#)) corrects the events that appear in the SEL3.

In addition to providing the most complete set of events, interactive analysis provides the standard by which automatic station and network processing are measured.

Preparing for Analysis

To prepare for analysis (process 2 in [Figure 32](#)), analysts quickly study the activity for the entire 24-hour period to determine if any large events have aftershocks, to determine the number of events that must be reviewed, and to note the stations that are not operational. Each analyst then reserves one of the time periods of data for analysis.

Selecting Events for Review

Events are not usually reviewed in sequential order, and are selected based on priorities (process 3 in [Figure 32](#)). Events that might affect other analysts events are analyzed first. This priority group includes large events with aftershocks (or potential for aftershocks) and events whose phases span the time periods that are being analyzed. After these events have been analyzed, the results are communicated to other analysts. The next priority is to analyze aftershocks from events that are not in the time period being analyzed. By scanning for these aftershocks, a large portion of the events in the analysis period may be reviewed quickly. Finally, the remaining events are reviewed.

Analyzing Events

Events in the SEL3 have locations in one of three environments: within the solid earth (terrestrial events), within an ocean (oceanic events), or within the atmosphere (atmospheric events). The environment of the event dictates the procedures used during analysis (processes 4, 5, and 6 in [Figure 32](#)). Techniques

available to analysts via the interactive ARS application include display manipulation (sorting, zooming, and other features), frequency filtering, beam forming, f-k analysis, polarization analysis, phase prediction, event location and depth estimation, spectral analysis, cepstral analysis, amplitude-period measurement, map display, and more. Additional techniques may be added as needed.

Terrestrial Events

Terrestrial events occur in the solid earth, including under the sea floor. The process for reviewing terrestrial events (process 4 in [Figure 32](#)) is described as follows:

1. Data are inspected visually to discover obvious inconsistencies.
2. The arrivals automatically associated with the event are evaluated. These arrivals may be re-timed, renamed, and their signal characteristics (azimuth, slowness, amplitude) reestimated.
3. A preliminary event location is made using only the best arrival information. If the location is poor, and an event cannot be built, the event is rejected. If the location is poor, but additional arrivals might be used to build the event, then different arrivals may be associated or reevaluated.
4. The location of the event is refined by associating additional phases, and the depth of the event is estimated by associating depth sensitive phases.
5. Secondary phases, hydroacoustic phases, and infrasonic phases are associated with the event (but are not used in estimating the location or magnitude). Associating these phases keeps them from contributing to other events.

Oceanic Events

Oceanic events occur in or immediately adjacent to the ocean. These events include underwater volcanic eruptions or man-made explosions that occur within the water column. In addition to hydroacoustic signals, such events may also generate seismic waves. The procedures for reviewing oceanic events (process 5 in

[Figure 32](#)) is similar to the review of terrestrial events, but are somewhat simpler due to the reduced number of valid phases that must be reviewed and the reduced number of events that are typically detected.

1. The data are first inspected visually to discover obvious inconsistencies.
2. Automatically detected arrivals are evaluated; signals are re-timed, phases are renamed, and parameters are reestimated as needed.
3. H phases are added to the events as needed, and a preliminary location is made. If the location is poor, the data are reevaluated to see if a valid event can be built. If not, the event is rejected.
4. If the location is accepted, then seismic and infrasonic phases are associated with the event whenever possible.

Atmospheric Events

An atmospheric event produces infrasonic waves and may also generate seismic or hydroacoustic waves if the event is close to the surface of the earth. The procedure for reviewing atmospheric events (process 6 in [Figure 32](#)) is similar to the review of terrestrial events, but is simpler due to the reduced number of valid phases that must be reviewed and the reduced number of events that are typically detected.

1. The data are first inspected visually to discover obvious inconsistencies.
2. Automatically detected arrivals are evaluated; signals are re-timed, phases are renamed, and parameters are reestimated as needed.
3. I phases are added to the events as needed, and a preliminary location is made. If the location is poor, the data are reevaluated to see if a valid event can be built. If not, the event is rejected.
4. If the location is good, then seismic and hydroacoustic phases are associated with the event. Such associations would be governed by the geographical location of the event; for example, if an atmospheric event was located above the interior of a continent, the event would not be searched for associated hydroacoustic phases.

Checking Event Quality

After review, each event is checked for adherence to the minimum standards for defining an event in the REB, for obvious mistakes in association, and for being typical of other events (process 7 in [Figure 32](#)). When a problem is discovered, the event is rechecked for mistakes.

The standards for defining an event in the REB, or event definition criteria, include a weighted sum of the contributions that each phase makes to the event solution and the types of phases that must be present for terrestrial events (see [“Defining Events” on page 153](#)). The event definition criteria are more stringent for the REB than they are for the event lists (see [“Applying Event Definition Criteria” on page 86](#)).

Association errors include events with associated hydroacoustic phases for which the path is clearly blocked, regional phases associated with events at teleseismic distances, and so on.

Each event is also checked to see if it is typical of the events usually observed. Events that are deep in areas where deep seismicity have not historically occurred are examples of anomalous events.

Scanning for Aftershocks

After analyzing a large event or any event that has generated aftershocks, the data are searched for aftershocks to minimize their impact on the analysis of subsequent events and later routine scanning (process 8 in [Figure 32](#)). The data from the main shock are displayed so that the first arrivals are aligned. The analyst scrolls the data across the computer display looking for phase alignments that indicate events in roughly the same location as the main shock.

Checking Analysis Quality

When all automatically determined events in the SEL3 for the time period being analyzed have been reviewed, the reviewed events are checked for completeness (process 9 in [Figure 32](#)). All pertinent signal and event characteristics should have been calculated. Incomplete events are completed before proceeding to the next step.

Scanning for Missed Events

The scan for missed events is a search for signals that the automated system did not associate to events (process 10 in [Figure 32](#)). The amount of data that is collected is too large to be scanned completely, so methods have been devised to reduce the amount of data that must be scanned without missing events. Missed events are often small events that are recorded only by stations within regional distances. Data from key stations that are sensitive to regional phases are the only data scanned. Scanning is performed regionally, using subsets of the key stations that cover the region being scanned. The data are examined for events that are missed, and then the potential events are evaluated in the same manner as the automated events. If data from non-key stations are needed, they can be added.

Checking Added Event Quality

Events added during the missed event scan are checked for adherence to the event definition criteria, for poor associations, and to see if the events are reasonable (as previously done for the automated events) (process 11 in [Figure 32](#)).

Checking Quality of Bulletin

After analysis has been completed, the Lead Analyst and experts in the fields of seismology, hydroacoustics, and infrasonics check the quality of all analyses that contribute to the REB (process 12 in [Figure 32](#)). Errors and inconsistencies in the phase associations, locations, and magnitudes are corrected either by the Lead Analyst or by the original analyst.

Records are kept of the changes made during interactive analysis so that the automated processes can be improved. Typical statistics include the numbers and percentages of all REB defining (used in the event solution) phases that are retimed, added, associated, renamed, and disassociated. The numbers and percentages of SEL3 events that were rejected and REB events that were repaired SEL3 split events, unmodified SEL3 events, or added during interactive analysis are also kept.

Corrections made by analysts contribute to changes in event location and depth. The location changes are usually presented as the numbers and percentages of REB events that changed by less than 10 km, between 10 and 50 km, and by more than 50 km. The changes are further subdivided by the distance to the nearest station (less than 200 km, between 200 and 2000 km, and more than 2000 km). Statistics on depth changes are also given by the distance to the nearest station.

DEFINING EVENTS

Minimum standards for the amount and type of information required for events included in the bulletins (for example, REB, SEB, and SSEB) are known as “event definition criteria.” Together with the quality of the sensor networks and the quality of processing at the IDC, the event definition criteria are critical in determining the number and reliability of events.

The event definition criteria can be adjusted to accommodate changes in the monitoring network or changes in requirements for bulletin reliability and sensitivity. Criteria that maximize the number of events also include more events that are invalid or have poorly determined hypocenters. More restrictive criteria reduce the number of events, but a higher proportion of these events are valid and have well constrained hypocenters.

The same event definition criteria are used for the REB, SEB, and SSEB and are based on the weighted count of defining phase observations from primary seismic, hydroacoustic, and infrasonic stations. [Table 15](#) shows the weights for different types of phases and observations that are used for the bulletins. The criteria currently applied are as follows:

- weighted count of 4.6 or greater
- for terrestrial events, first arriving P-type phases at three Primary seismic stations

With this weighted count threshold and the weights given in [Table 15](#), the minimum requirements for events in each medium are as follows:

- terrestrial–P-type phase measurements (time azimuth and slowness) from two primary seismic arrays and time only for a primary 3-C station,
- oceanic–H phase times recorded at three hydroacoustic stations
- atmospheric–I phase arrival times and azimuths at two infrasonic stations, and I phase azimuth only at a third infrasonic station

TABLE 15: PHASE WEIGHTS FOR EVENT DEFINITION IN REB, SEB, AND SSEB

Phase Type	Station Type	Onset Time	Azimuth	Slowness
P-type primary	array	1.00	0.40	0.40
P-type secondary	array	0.40	0.00	0.00
S-type regional	array	0.70	0.40	0.40
S-type teleseismic	array	0.70	0.00	0.00
P-type	3-C	1.00	0.20	0.20
P-type secondary	3-C	0.40	0.00	0.00
S-type regional	3-C	0.70	0.00	0.00
S-type teleseismic	3-C	0.70	0.00	0.00
H-type	hydroacoustic	1.54	0.00	0.00
I-type	infrasonic	0.8	1.0	

TECHNIQUES

A variety of techniques have been developed to assist and improve analysis. These techniques, which are used to enhance signals, identify phases, recognize and reject detected signals that are natural or artificial noise, and to categorize and evaluate events, are described in [\[IDC6.2.5\]](#).

Enhancing Signals

Filtering is the most common technique for improving the appearance of signals to aid their identification and subsequent analysis. Analysts can select from a wide range of bandpass, high-pass, or low-pass filters. The optimum filter depends upon the type of phase, distance to the source, the underlying instrument response, and the signal and noise characteristics of each particular recording station.

The waveform display can be scaled both vertically ("rescaling") and horizontally ("zooming") to ease analysis. The default automatic vertical scaling may be inappropriate where noise "spikes" or exceptionally large signals are present. Zooming is used to put the particular signal being analyzed in the best noise context.

Data from sensor sites of an array are commonly combined into "beams" (see ["Beam Generation" on page 191](#)) to enhance snr. The ARS application allows the analyst to form beams in any desired direction, with the most common selection being based upon the results of f-k processing (see the following paragraphs and ["Frequency-wavenumber Analysis" on page 206](#)). Beams can help confirm the grouping of phases, and a poor beam wave can indicate that an event is greatly mislocated.

Frequency-wavenumber (f-k) analysis can improve the estimates of azimuth and slowness that were generated by the automatic processing and help confirm the association of phases added during analysis. This technique also assists in identifying noise detections, incorrectly labeled phases, and the separation of mixed events. Spectral analysis, using the ARS tool *Spectraplot* is valuable for identifying the frequency band in which signal power exceeds noise power.

Identifying Phases

Phases are identified by signal characteristics such as frequency content, impulsiveness, duration, amplitude, slowness, azimuth, and so on. A detailed description is given in “Identifying Phases” on page 53 in [\[IDC6.2.5\]](#).

Seismic phases are typically identified by their context within a wave train and their arrival times. Confirmation is often provided by slowness and azimuth, and frequency content and duration are also helpful. Depth phases are best identified through the “moveout” of such signals with increasing distance as observed at a number of stations for a particular event. For large events in particular, as many phases as possible should be identified so that false events created from prominent late-arriving secondary phases such as PKKP or PKPPKP can be recognized.

Experience with the identification of hydroacoustic and infrasonic phases is significantly less than for seismic phases. Signal envelope and frequency content are the most valuable characteristics. The earlier chapters [“Hydroacoustic Station Processing” on page 43](#) and [“Infrasonic Station Processing” on page 61](#) describe the characteristics that are most commonly used to identify the various types of phases.

Noise and Data Irregularities

Detections due to noise and data irregularities must be recognized and eliminated from further processing. Normal natural background noise is the easiest noise to recognize, and much of it originates from oceanic waves coupled into the earth as microseisms and into the atmosphere as microbaroms. Biological noise is a common problem in hydroacoustic data, with whales being the biggest culprits.

Cultural noise is created by man-made sources such as quarrying operations, trains, ships, and aircraft. Such noise is often relatively monochromatic and long in duration. The careful siting of most stations is intended to minimize both natural and cultural noise, but political (influence on site selection) or logistical (availability of power/land on which to locate seismic or infrasound sensors in broad oceanic areas) considerations have in many cases dictated less desirable locations for some sites.

Hardware problems may create electrical or mechanical noise within instrumentation. The appearance of this in the data may be affected by filters in the recording systems so that, for example, simple spikes may be broadened and made to look more complex. A similar effect may be created when data dropouts caused by communications problems are subjected to filters during either the automatic processing or interactive analysis.

Categorizing and Evaluating Events

In addition to the major categories of terrestrial, oceanic, and atmospheric events, some events may belong to several subcategories.

Terrestrial events are categorized by depth (deep or shallow), size (small to large) and location (continental or beneath the ocean). Deep events are often readily recognized by the types of phases observed in conjunction with the typical impulsive character. Shallow events tend to generate signals with more emergent onsets and longer durations.

During bulletin analysis and scanning, events are evaluated by a variety of techniques that include fixing location to a common source region, fixing depth, testing particular hypotheses by temporarily adding or renaming phases, and recalculating f-k. With experience, the analyst recognizes locally consistent groups of stations based on closeness and background noise levels at each station. "Hot" stations, which are sensitive to activity within a given source region, are available through an ARS tool.

Post-analysis Processing

This chapter describes post-analysis IDC processing and includes the following topics:

- [Overview](#)
- [Estimating Event Characteristics](#)
- [Applying Standard Event Screening Criteria](#)

Post-analysis Processing

OVERVIEW

Post-analysis processing begins with supplementary automatic processing to calculate event-based characteristics that require the accurate estimates of the arrival times, phase names, and event locations provided through interactive analysis. Because events are added during analysis and new signals are created/associated with events, some signal features and characteristics (for example, snr, amplitude, and period) are also computed in this stage. The events and their event characteristics are published in the SEB.

Post-analysis processing continues by using the event characteristics to screen out events that are natural or non-nuclear. The event characteristics are compared to event screening criteria. The events in the SEB that pass the event screen are published in the SSEB.

The Executive Summary includes information such as the number of events detected, the numbers of events in the various screening categories, and the status of the IMS network, the communications links, and the IDC.

[Figure 33](#) shows the steps of post-analysis processing, and [Table 16](#) lists the software modules that support each step. Although the processes are numbered, the tasks may occur in any order.

▼ Post-analysis Processing

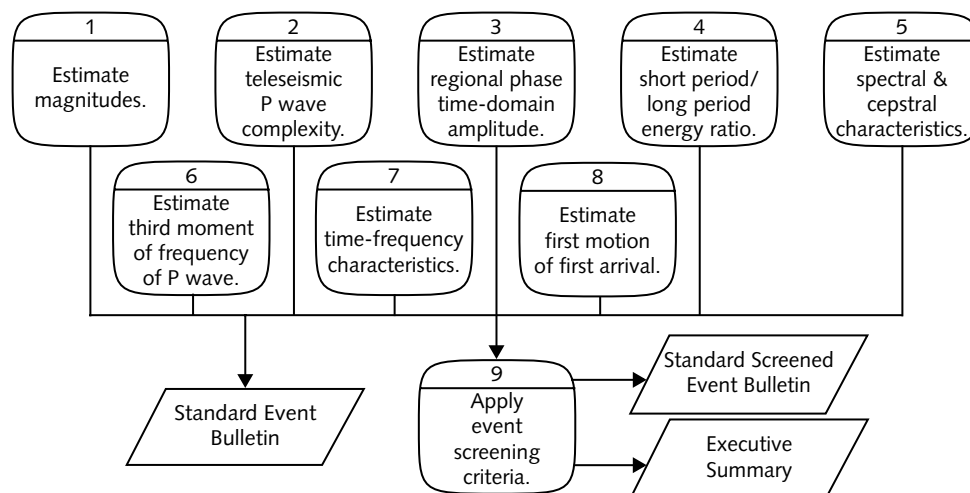


FIGURE 33. POST-ANALYSIS PROCESSING

TABLE 16: SOFTWARE USED IN POST-ANALYSIS PROCESSING

Step	Software	Reference(s)
1 Estimate magnitudes.	DFX, Maxsurf, EvLoc	[Wah96a] , [Wah96b] , [Ste97]
2 Estimate teleseismic P wave complexity.	DFX	[Wah96a] , [Wah96b]
3 Estimate regional phase time-domain amplitude.		
4 Estimate short-period/long-period energy ratio.		
5 Estimate spectral and cepstral characteristics.		
6 Estimate third moment of frequency of P wave.		
7 Estimate time-frequency characteristics.		
8 Estimate first motion of first arrivals.		
9 Apply event screening criteria.	EvScreen	

ESTIMATING EVENT CHARACTERISTICS

Event characteristics calculated during post-analysis processing require the accurate phase information and event locations estimated during interactive analysis as well as the signal characteristics estimated during station processing. If phases have been added during analysis, the signal characteristics needed to establish values for the event characteristics are estimated.

Estimating Magnitudes

During network processing the body wave magnitude, m_b , and local magnitude, M_L , were calculated based on amplitude measurements made during station processing. In post-analysis processing, the event location is improved from interactive analysis allowing a more accurate estimate of these magnitudes. In addition, surface wave magnitudes M_s , and maximum-likelihood m_b and M_s are also calculated during post-analysis processing.

Maximum-likelihood estimates use amplitude measurements also from stations that did not detect the signals. When signals are not observable at a station, the amplitude of the noise at the predicted time of the arrival is used.

Body Wave Magnitude

Arrivals are often added or changed during interactive analysis. This process can invalidate the amplitudes and periods measured during station processing. Before the magnitude is estimated again, the amplitude and period are measured again based on new arrival times (or the predicted arrival times when no signal is observed). For observed phases, amplitude and period are measured in the same way that they were for station processing (see [“Measuring Amplitude and Period” on page 27](#)). For stations between 20 and 100 degrees from the event that recorded no P phase, the amplitude and period are measured at the predicted phase arrival time. The time window used for measuring noise amplitude and

period is larger (starting 3.0 seconds prior to the predicted arrival time and ending 9.0 seconds after that time) than the standard time window to account for the uncertainty in the calculated arrival time.

Station magnitudes and network average magnitudes are recalculated using the new amplitude and period measurements of the arrivals, and are stored in the **stamag** and **netmag** database tables (*amptype* = A5/2) as *magnitude* with *magtype* = mb_ave. The uncertainty of a network average magnitude in the database table **netmag** is calculated as the standard deviation of the network magnitude. In addition, maximum likelihood estimates (MLE) of magnitudes are calculated. The MLE magnitudes are based on amplitude and period measurements of both arrivals (*amptype* A5/2) and noise (*amptype* ANP/2) and are also stored in the **stamag** and **netmag** tables as *magnitude* with *magtype* = mb_mle. The MLE calculation follows [Rin76]; for a given event the observations (based on both signal and noise) are assumed to be Gaussian with an unknown mean value, which is estimated, and have a standard deviation of 0.35. The estimates are currently based on data from primary stations. The uncertainty of MLE network magnitudes are obtained with bootstrapping according to a procedure described by McLaughlin [McL88].

Details of m_b magnitude estimation are provided in [“Estimating Body Wave Magnitudes” on page 234](#).

Local Magnitude

Amplitudes for M_L estimation are measured in the same manner as used for station processing (see [“Measuring Amplitude and Period” on page 27](#)). M_L is then recalculated as it was during network processing (see [“Estimating Local Magnitudes” on page 237](#)); individual station magnitudes, based on arrivals with *amptype* = SBSNR, are stored in the **stamag** database table, and their weighted averages are stored as network magnitudes in the **netmag** database table as *magnitude*. The *magtype* = mlppn is used for M_L magnitudes in the **stamag** and **netmag** tables. No maximum likelihood estimate is made for M_L .

Surface Wave Magnitude

Amplitudes and periods are calculated automatically as part of surface-wave processing after events have been formed and accepted during interactive analyst review.

The measurements are made on a beam or single channel of vertical component data. Channels are prefiltered by replacing the instrument with a common response, that of a KS36000 instrument. The beam for arrays is formed by using the azimuth calculated to the origin and a phase velocity fixed at 3.5 km/s.

Amplitude (in nm) and period measurements for detected Rayleigh wave (LR) signals and noise are stored in the database table **amplitude** (defined by the attribute *amptype* = *ALR*/2 for signals and *amptype* = *ANL*/2 for noise). The amplitude (half peak-to-peak) of a wave train for noise or signal is selected from those measurements with periods in a given period range (18–22 seconds) that maximizes the station magnitude, which is the same as maximizing amplitude/period. The arrival time of the amplitude also has to be within the time window predicted for the period limits 18 and 22 seconds. The resolution of amplitude and period measurements is limited by the digital quantization and sampling rate.

Station and network M_s are computed only for primary seismic stations. The station and network magnitudes from signals are stored in the database tables **stamag** and **netmag** (defined by the attribute *magtype* = *ms_ave*). The uncertainty of a network average magnitude in the database table, **netmag**, is calculated as the standard deviation of the network magnitude.

In addition to average network magnitudes, MLE are estimated (*magtype* = *ms_mle*). The estimates are based on data from primary seismic stations and assumed standard deviations of station magnitudes of 0.25 magnitude units. The uncertainty of a MLE network M_s is, as for the MLE of m_b , obtained with bootstrapping. In cases where no signal amplitudes have been recorded at stations between 2 and 100 degrees, upper-bound magnitudes are calculated as the magnitude at which the probability of there being no detecting stations between 2 and 100 degrees is 0.95, again, assuming a Gaussian distribution.

Details of M_s estimation are provided in [“Estimating Surface Wave Magnitudes” on page 239](#).

Estimating Teleseismic P-wave Complexity

The measure of complexity (process 2 of [Figure 33](#)) follows the energy-ratio technique of Douglas [\[Efr79\]](#). In this method the coherent beam of the P signal is filtered in the band from 0.8 to 2.0 Hz, squared, and smoothed with an exponential window (with a time constant of 1.5s). Complexity is then computed as the ratio of the area under this curve (S) for a coda window following the P-signal window to the area for the P-signal window, both corrected for background noise (N). The default values for the signal and coda window lengths are 5 and 30 seconds, respectively. Complexity is thus defined as follows:

$$complexity = \frac{S(5, 35) - N}{S(0, 5) - N} \quad (35)$$

with the exception of the following two cases:

- If $S(0, 5) - N < 0$, complexity is undefined and is set to -1 ;
- If $S(5, 35) - N < 0$, all energy is in the first 5s, so complexity is set to 0.0.

Complexity is measured only for stations that are associated to an event and are at teleseismic distances (30–90 degrees). If a PcP phase is observed within 35 seconds of P, no measurement is made. The complexity measurements are stored in the **complexity** table of the IDC database [\[IDC5.1.1Rev1\]](#).

Estimating Regional Phase Time-domain Amplitudes

The *DFX* application can compute time-domain amplitude measurements on a variety of beams for specified time windows. Amplitude measurement types include absolute maximum amplitude, maximum peak-to-trough, maximum peak-to-peak, root-mean-squared (rms) amplitude, and so on. Beam types include coherent, steered, incoherent, and rms measured on vertical, radial, or transverse

Post-analysis ▼ Processing

components. Time intervals for the measurements can be based on predicted travel times of seismic phases or on fixed group velocity windows, or they may be computed from observed arrival times in the database.

The following time-domain regional-phase amplitude measurements are computed for each seismic station that is associated to and within 20 degrees of an analyst-reviewed event (process 3 of [Figure 33](#)):

- absolute maximum amplitude on 2–4, 4–6, 6–8, 8–10, 10–12, 12–14 Hz rms beams for standard signal and noise windows around *observed* Pn, Pg, Sn, and Lg phases.

The signal and noise windows are:

signal: 5 seconds before to 15 seconds after onset

noise: 15 seconds before to 5 seconds before onset

- absolute maximum amplitude on 2–4, 4–6, 6–8, 8–10, 10–12, 12–14 Hz rms beams for *predicted* time/velocity windows around Pn, Pg, Sn, Lg phases:

Pn: 8s before the theoretical arrival time of Pn to a group velocity of 6.4 km/s

Pg: 6.3 to 5.8 km/s group velocity

Sn: 5s before the theoretical arrival time of Sn and a 20s duration

Lg: 3.7 to 3.0 km/s group velocity

- absolute maximum amplitude on 2–4, 4–6, 6–8, 8–10, 10–12, 12–14 Hz rms beams for *predicted* time and/or velocity windows around pre-Pn, pre-Pg, pre-Sn, and pre-Lg noise:

Pre-Pn: 13s before the theoretical arrival time of Pn and a 5s duration

Pre-Pg: 6.4 to 6.3 km/s group velocity

Pre-Sn: 10s before the theoretical arrival time of Sn and a 5s duration

Pre-Lg: 3.8 to 3.7 km/s group velocity

The windows for the crustal modes (Pg and Lg as well as Pre-Pg and Pre-Lg) are defined in terms of the well-defined group velocity bounds that characterize such phases. In contrast, the windows for the body waves (Pn and Sn as well as Pre-Pn and Pre-Sn) are defined (with a single exception) in terms of the arrival time, which delineates such phases with less ambiguity.

The instrument response correction at the band center frequency is applied to all of the amplitude measurements. Also, amplitudes in the higher frequency bands are only calculated for stations with appropriate sampling rates. These signal and noise amplitude measurements are stored in the **amplitude** table of the database.

Estimating Energy Ratio

The short-period/long-period energy ratio event-characterization parameter is the ratio of the integrated energy between the short-period (SP) vertical component P-wave train and the sum of the three long-period (LP) surface wave trains (process 4 of [Figure 33](#)) [\[Sai96\]](#). The calculation of the energy ratio is restricted to 3-C broadband stations within 15 degrees of the event. The P wave train is the filtered (1–8 Hz) time series between a group velocity of 8.2 km/s and 4.5 km/s, and the LP surface wave trains are the filtered (0.033–0.167 Hz) Z, N and E channel time series between a group velocity of 4.5 km/s and 2.5 km/s. SP/LP energy ratio values can be found in database table **splp**.

Estimating Spectral and Cepstral Characteristics

Each associated regional phase within 20 degrees is analyzed for spectral and cepstral characteristics [\[Bau88\]](#). The measurements from this analysis are defined in the following list and stored in the database (process 5 of [Figure 33](#)):

- The variance of the detrended log spectrum of each associated regional phase is calculated for frequency bands with width ≥ 4.0 Hz and with $\text{snr} \geq 3.0$. The variance and the coefficients of the quadratic trend, removed from the log spectrum, are stored in table **spvar**.

Post-analysis ▼ Processing

- Amplitudes and quefrequencies of all cepstral peaks with amplitudes above a default threshold of 0.032 that are not present in the preceding noise window are calculated. A 5.0 s window starting 0.3 s before the time of the associated phase is used for the signal, and the preceding 5.0 s window is used for the noise. Cepstral peaks are only declared if they are consistent (have the same quefrequency) among two or more phases at the same station. The maximum amplitude of the individual phase peaks for that station and the quefrequency are stored in database table **ceppks**.

Estimating Third Moment of Frequency

The source spectrum of the various source types vary to some degree, with nuclear explosion sources having higher frequency content than equivalently sized earthquakes. The third-moment-of-frequency (*tmf*) measure is employed to find these differences and to assist with event screening (process 6 of [Figure 33](#)).

The *tmf* measure follows that of [\[Wei71\]](#) and is calculated as follows:

$$tmf = \left(\frac{\int_{f_{min}}^{f_{max}} f^3 sig(f) df}{\int_{f_{min}}^{f_{max}} sig(f) df} \right)^{\frac{1}{3}} \quad (36)$$

where the signal has been corrected for noise:

$$sig(f) = signal(f) - 0.5(N_1(f) + N_2(f)) \quad (37)$$

$N_1(f)$ and $N_2(f)$ are two distinct noise spectra made before the onset of the signal.

This measure gives more weight to high frequencies, so events with a greater proportion of high frequency energy will have higher *tmf* values. Typically f_{min} and f_{max} are set to 0.5 and 5.0 Hz, respectively. All associated P phases greater than 20 degrees are measured of *tmf*. However, only those measurements for which the signal spectrum amplitudes are greater than corresponding noise spectrum ampli-

tudes for more than a specified percentage of the frequencies (default, 90%) are stored in the **thirdmom** database table. The latter restriction is to provide some quality control of the measurements.

Estimating Time-frequency Characteristics

Waveforms produced by multiple source events (ripple-fired mining blasts) commonly show spectral modulations, and often the modulations extend well into the coda. Time-frequency parameters (2-D cepstral peaks and correlation measurements) are calculated using the multi-taper sonogram technique to detect these delayed-fired diagnostics (see process 7 of [Figure 33](#), [\[Hed89\]](#), and [\[Hed90\]](#)).

The calculation involves converting the time series (first 30s of coda) into a time-frequency (t-f) matrix. After converting the matrix to a binary form that includes randomizing all points that fall below the noise level, an estimate of how well the t-f matrix is correlated in time is calculated. Also, the maximum amplitude and associated quefrequency of the 2-D Fourier transform of the binary t-f matrix is determined, and the zero-lag cross correlation between the three pairs of binary t-f matrices (for 3-C stations only) is estimated. Data from all associated stations less than 15 degrees from an event are measured for t-f characteristics, which are stored in the **timefreq** database table.

Estimating First Motion

First motion is calculated for the first arriving phase of each associated waveform (process 8 of [Figure 33](#)) and is restricted to phases Pn, Pg, P, and PKP. First motion is determined from the filtered (0.5–5.0 Hz) beam by finding the first maximum or minimum amplitude after the measured onset time of the arrival. The signed amplitude of this turning point is returned as the measure of first motion and is stored in the **amplitude** database table. This parameter is still in the early stages of development and testing. The rate of success of the first motion measurement must be improved by either improving onset picks or by improving the sophistication of the algorithm.

APPLYING STANDARD EVENT SCREENING CRITERIA

Seismic event-screening criteria, based on depth and $M_s:m_b$, are applied on a routine basis to the events in the SEB. Event location is also used to categorize events based on location error ellipses and whether the event is onshore or offshore, but location is not explicitly used to screen out events. This section describes the initial version of the standard screening criteria (process 9 of [Figure 33](#)). Changes to these provisional screening criteria may be made in the future based on recommendations by Working Group B of the CTBT Preparatory Commission.

Screening Events by Depth

Given the depth estimate of an event, *origin.depth*, and the variance of the depth estimate, *origerr.szz*, the event is “screened out” for having a depth greater than 10 km (default) at the 97.5% confidence level if

$$depth - 2\sigma_D > 10 \quad (38)$$

where $2\sigma_D = 2\sqrt{szz} + k$.

The term, k , accounts for errors in the depth estimate due to regional biases that are not adequately represented by szz . The value of k depends on whether or not the following depth criteria are met:

- At least three depth phases are detected for an event (*origin.ndp* ≥ 3); and
- the measured moveout of pP-P arrival times from the nearest to the farthest stations is at least 1.5 seconds.

If these depth phase criteria are satisfied, a value of $k = 0$ is used. For all other unconstrained depth estimates a conservative value of k is used. k is regionally dependent and is significantly smaller for regions that have been calibrated. In uncalibrated regions, $k = 40$ km. For further details, see [“Depth” on page 243](#).

A “score” is computed to indicate numerically for each event with an unconstrained depth estimate the degree to which that event meets or does not meet the event-screening criteria:

$$\text{score}_{\text{depth}} = \frac{\text{depth} - 10}{2\sigma_D} - 1 \quad (39)$$

A positive score indicates that the depth is greater than 10 km at the 97.5% confidence level. Because a 97.5% confidence level is used, the score increments one unit for each two-sigma interval from the depth screening threshold of 10 km.

Screening Events by $M_s:m_b$

Screening based on M_s and m_b uses the difference between the network magnitudes, $Am_b - M_s$, where A is a constant slope term to account for magnitude dependence of the $M_s:m_b$ relation. Assuming that $Am_b - M_s$ has a normal distribution, an event is screened out at the 97.5% confidence level for having a magnitude difference ($Am_b - M_s$) less than a threshold, M_T , if

$$Am_b - M_s + 2\sigma_M > M_T \quad (40)$$

where σ_M is the standard deviation of $Am_b - M_s$ (see [“Surface Wave Magnitude:Body Wave Magnitude” on page 245](#)). Based on calibration of PIDC magnitudes relative to historical earthquake and explosion data, the standard default values of $A = 1.25$ and $M_T = 2.20$ are used [\[Mur97\]](#).

A score is computed to indicate numerically for each event with M_s and m_b the degree to which that event does, or does not, meet the $M_s:m_b$ screening criterion:

$$\text{score}_{\text{mbms}} = \frac{2.20 - 1.25m_b + M_s}{2\sigma_M} - 1 \quad (41)$$

A positive score indicates that $1.25m_b - M_s$ is less than 2.20 at the 97.5% confidence level. Because a 97.5% confidence level is used, the score increments one unit for each two-sigma interval from the screening threshold of 2.20.

Combining Screening Scores

A combined screening score is computed to indicate numerically the degree to which an event does or does not meet the screening criteria. The combined score based on depth and $M_s:m_b$ screening criteria is given by the following equation:

$$\text{score}_{\text{combined}} = \frac{1}{\sqrt{2}} \left[\frac{\text{depth} - 10}{2\sigma_D} + \frac{2.20 - 1.25m_b + M_s}{2\sigma_M} \right] - 1. \quad (42)$$

The factor of $\sqrt{2}$ normalizes the combined score such that a score of zero corresponds to a 97.5% confidence level such that an explosion at a maximum depth of 10 km would not be screened out (see ["Combined Screening Score" on page 246](#)).

For cases in which only depth or $M_s:m_b$ is available for an event, the combined score is given by the appropriate individual score.

Screening Category Definitions

The initial version of the standard event-screening criteria and definitions of the event-screening categories are described in this section.

[Table 17](#) defines the screening categories and criteria, based on depth and $M_s:m_b$. Events below m_b 3.5 are not considered for application of the screening analyses, because the screening criteria are typically applicable only to events that are large enough that reliable depth and m_b – M_s measurements can be obtained, and because the existing screening criteria do not include capabilities to screen out chemical mining blasts, which are usually smaller than m_b 3.5.

Events for which M_s and unconstrained depth estimates are not available are categorized as "Insufficient Data," and are considered to be not screened out because neither the depth nor $M_s:m_b$ screening criteria could be applied.

An event above m_b 3.5 with depth/ $M_s:m_b$ measurements is categorized as "Not Screened Out" or "Screened Out" depending on whether the combined score is less than or equal to zero or greater than zero, respectively.

TABLE 17: SEISMIC EVENT-SCREENING CATEGORIES AND CRITERIA

Category	Criteria
Not Considered	$m_b < 3.5$
Insufficient Data	depth constrained to the surface and no M_s
Not Screened Out	$\text{score}_{\text{combined}} \leq 0$
Screened Out	$\text{score}_{\text{combined}} > 0$

Categorizing Events by Location

Event location is not used in the standard event-screening software to explicitly screen out events that are offshore. However, location error ellipses obtained from the hypocentral inversion are used to separate events into three subcategories: Offshore, Onshore, and Mixed. An onshore/offshore grid (including islands) with 5-minute resolution is currently used for this determination. The fraction of offshore grid cells, *Foffshore*, enclosed by or touching the 90% location error ellipse, is computed for all events and provided in the SEB and SSEB. [Table 18](#) defines the location categories in terms of *Foffshore*.

TABLE 18: DEFINITIONS OF LOCATION CATEGORIES

Location Category	Criterion	Definition
Offshore	$F_{\text{offshore}} = 1.0$	90% location error ellipse contains or touches only offshore cells
Onshore	$F_{\text{offshore}} = 0.0$	90% location error ellipse contains or touches only onshore cells
Mixed	$0.0 < F_{\text{offshore}} < 1.0$	90% location error ellipse contains or touches onshore and offshore cells

Each event in the SEB is assigned to one of the four screening categories in [Table 17](#), as well as to one of the three location categories in [Table 18](#).

Algorithms

This chapter describes the algorithms used in the processing and analysis functions of the IDC and includes the following topics:

- [Overview](#)
- [Frequency Filters](#)
- [Interpolation](#)
- [Beam Generation](#)
- [Normalized Cross-correlation Functions](#)
- [F-statistics](#)
- [STA/LTA Detectors](#)
- [Seismic Arrival-time Estimation via Akaike Information Criterion](#)
- [Frequency-wavenumber Analysis](#)
- [Modified Cramer-Rao Variance Estimation](#)
- [Polarization Analysis](#)
- [Cepstral Analysis](#)
- [Noise Spectrum Equalization](#)
- [Probability Weighted Time](#)
- [Neural Networks](#)
- [Bayesian Inference](#)
- [Detection Probability](#)
- [Event Screening Statistics](#)

Algorithms ▼

- [Automatic Association Chi-Squared Test](#)
- [Event Location Error Estimation](#)

Algorithms

OVERVIEW

The algorithms included in this chapter supplement the processing descriptions of the previous chapters.

FREQUENCY FILTERS

Infinite Impulse Response Filters

The principal mechanism for spectrally filtering digital waveform data is a package (implemented in *FORTRAN* and linked to C), which designs and applies infinite impulse response (IIR) filters, also referred to as recursive filters. The following paragraphs describe the design and application of these filters.

Design

The IIR filter design algorithm is capable of generating IIR filters of the following types: lowpass, highpass, bandpass, and band reject—any type with up to 10 poles per filter. The design is based on the application of a cascade of filter sections, where each section has an identical second-order architecture. The number of sections is governed by the number of poles, N_p , and the filter type. Where the filter type is either lowpass or highpass and N_p is odd-valued, the number of sections, N_s , is $(N_p+1)/2$. Where the filter type is either lowpass or highpass and N_p is even-valued, N_s is $N_p/2$. Where the filter type is either bandpass or band reject, the number of sections, N_s , is N_p . Generally, IIR filter response functions attenuate out of band energy at 6 dB per octave from the band edge, per pole; therefore, a 5-pole filter will attenuate at 30 dB per octave. The IIR filter design process consists of the following steps:

Algorithms ▼

1. Generate a set of equispaced, complex-valued poles in the left-half of the complex-plane in accordance with the number of poles specified by the user. Where the number of poles is odd valued, the first pole lies on the negative real axis. The poles are generated in the upper-half of the left-hand side of the complex-plane because the IIR filters designed by this algorithm are characterized by conjugate pair of poles. Where N_p is even valued, the initial value of the n^{th} pole (n counting from 1 to $N_p/2$) is given by equation (43):

$$p[n] = e^{-\pi\sqrt{1} \cdot \left\{ \frac{1}{2} + [(2n-1)/2N_p] \right\}} \quad (43)$$

Where N_p is odd-valued, $p[1] = -1$, a real-valued number; and for odd-valued $N_p > 1$, the other poles (n counting from 2 to $[(N_p-1)/2]$) are given by equation (44):

$$p[n] = e^{-\pi\sqrt{1} \cdot \left\{ \frac{1}{2} + [(2n-3)/2N_p] \right\}} \quad (44)$$

2. Transform the user-requested filter frequency limit(s), f , via the tangent transformation (as shown in equation (45)), which adjusts for the analog bilinear transformation of an “analog” continuum time filter to a “digital” discrete time filter

$$\tilde{f} = \frac{1}{2\pi} \cdot \tan \left[\pi \cdot \left(\frac{f}{F_s} \right) \right] \quad (45)$$

where the sampling frequency is F_s .

3. Generate second-order section coefficients from an all-pole description. This step must be done in one of four distinct ways depending on the type of IIR filter: lowpass, highpass, bandpass, or band reject.

For lowpass or highpass filters, the numerator ($sn[k]$) and denominator ($sd[k]$) coefficients are generated according to equations (46–51).

For the n^{th} pole in the lowpass case, regardless of whether N_p is odd or even valued, the numerator coefficients are as follows:

$$\begin{aligned} \text{sn}[3(n-1)+1] &= 1 \\ \text{sn}[3(n-1)+2] &= \text{sn}[3(n-1)+3] = 0 \end{aligned} \quad (46)$$

or for the highpass case, use equations (47):

$$\begin{aligned} \text{sn}[3(n-1)+1] &= \text{sn}[3(n-1)+2] = 0 \\ \text{sn}[3(n-1)+3] &= 1 \end{aligned} \quad (47)$$

Where N_p is even valued for the n th pole (n counting from 1 to N_p) or where N_p is odd valued and larger than 1 for the n th pole (n counting from 2 to N_p), for the lowpass case, use equations (48):

$$\begin{aligned} \text{sd}[3(n-1)+1] &= |p[n]|^2 \\ \text{sd}[3(n-1)+2] &= -2\text{Re}\{p[n]\} \\ \text{sd}[3(n-1)+3] &= 1 \end{aligned} \quad (48)$$

or for the highpass case, use equations (49):

$$\begin{aligned} \text{sd}[3(n-1)+1] &= 1 \\ \text{sd}[3(n-1)+2] &= -2\text{Re}\{p[n]\} \\ \text{sd}[3(n-1)+3] &= |p[n]|^2 \end{aligned} \quad (49)$$

And for the first pole, where N_p is odd-valued, for the lowpass case, use equations (50):

$$\begin{aligned} \text{sd}[1] &= -\text{Re}\{p[1]\} \\ \text{sd}[2] &= 1 \\ \text{sd}[3] &= 0 \end{aligned} \quad (50)$$

or for the highpass case, use equations (51):

$$\begin{aligned} \text{sd}[1] &= 1 \\ \text{sd}[2] &= -\text{Re}\{p[1]\} \\ \text{sd}[3] &= 0 \end{aligned} \quad (51)$$

For bandpass or band reject, the coefficients are generated according to the following scheme. For the n th pole, regardless of whether N_p is odd or even valued, for the bandpass case, use equations (52):

Algorithms ▼

$$\begin{aligned} \text{sn}[3(n-1)+1] &= \text{sn}[3(n-1)+3] = 0 \\ \text{sn}[3(n-1)+2] &= 2\pi \cdot (\tilde{f}_h - \tilde{f}_l) \end{aligned} \quad (52)$$

or for the band reject case, use equations (53):

$$\begin{aligned} \text{sn}[3(n-1)+1] &= 2\pi \cdot (\tilde{f}_h - \tilde{f}_l) \\ \text{sn}[3(n-1)+2] &= 0 \\ \text{sn}[3(n-1)+3] &= 1 \end{aligned} \quad (53)$$

Where N_p is even valued for the n th pole (n counting from 1 to N_p) or where N_p is odd valued and larger than 1 for the n th pole (n counting from 2 to N_p), for the bandpass case, use equation (54):

$$q[n] = p[n] \quad (54)$$

or for the band reject case, use equation (55):

$$q[n] = \frac{p^*[n]}{|p[n]|^2} \quad (\text{in other words, the inverse of } p[n]) \quad (55)$$

$$\text{sd}[6(n-1)+1] = \pi^2 |d|^2 \quad (56)$$

where $|d|^2$ is given by equations (57).

$$\begin{aligned} |d|^2 &= \left[q[n]f_{\text{diff}} + \sqrt{q^2[n]f_{\text{diff}}^2 - 4\tilde{f}_h\tilde{f}_l} \right] \cdot \left[q^*[n]f_{\text{diff}} + \sqrt{q^{*2}[n]f_{\text{diff}}^2 - 4\tilde{f}_h\tilde{f}_l} \right] \\ f_{\text{diff}} &= \tilde{f}_h - \tilde{f}_l \end{aligned} \quad (57)$$

$$\text{sd}[6(n-1)+2] = -\text{Re} \left\{ (2\pi) \cdot \left[q[n](\tilde{f}_h - \tilde{f}_l) + \sqrt{q^2[n](\tilde{f}_h - \tilde{f}_l)^2 - 4\tilde{f}_h\tilde{f}_l} \right] \right\} \quad (58)$$

$$\text{sd}[6(n-1)+3] = 1 \quad (59)$$

$$\text{sd}[6(n-1)+4] = \pi^2 |e|^2 \quad (60)$$

where $|e|^2$ is given by equations (61).

$$|e|^2 = \left[q[n]f_{\text{diff}} - \sqrt{q^2[n]f_{\text{diff}}^2 - 4\tilde{f}_h\tilde{f}_l} \right] \cdot \left[q^*[n]f_{\text{diff}} - \sqrt{q^{*2}[n]f_{\text{diff}}^2 - 4\tilde{f}_h\tilde{f}_l} \right] \quad (61)$$

$$f_{\text{diff}} = \tilde{f}_h - \tilde{f}_l$$

$$sd[6(n-1)+5] = -\text{Re} \left\{ (2\pi) \cdot \left[q[n](\tilde{f}_h - \tilde{f}_l) - \sqrt{q^2[n](\tilde{f}_h - \tilde{f}_l)^2 - 4\tilde{f}_h\tilde{f}_l} \right] \right\} \quad (62)$$

$$sd[6(n-1)+6] = 1 \quad (63)$$

And for the first pole, where N_p is odd-valued, for the bandpass case, use equations (64):

$$\begin{aligned} sd[1] &= (2\pi)^2 \cdot \tilde{f}_l \cdot \tilde{f}_h \\ sd[2] &= -2\pi(\tilde{f}_h - \tilde{f}_l) \cdot \text{Re}\{p[1]\} \\ sd[3] &= 1 \end{aligned} \quad (64)$$

or for the band reject case, use equations (65):

$$\begin{aligned} sd[1] &= -(2\pi)\tilde{f}_h\tilde{f}_l \cdot \text{Re}\{p[1]\} \\ sd[2] &= 2\pi(\tilde{f}_h - \tilde{f}_l) \\ sd[3] &= -\text{Re}\{p[1]\} \end{aligned} \quad (65)$$

4. In the case of lowpass and highpass filters only, adjust the coefficients, which govern the cutoffs, of lowpass or highpass second-order sections in accordance with a polynomial transformation. For the n^{th} filter section (n counting from 1 to N_s), the adjustment to the coefficients is accomplished via equations (66):

$$\begin{aligned} sn[3n-1] &\leftarrow \{sn[3n-1]/2\pi\tilde{f}\} \\ sd[3n-1] &\leftarrow \{sd[3n-1]/2\pi\tilde{f}\} \\ sn[3n] &\leftarrow \{sn[3n]/(2\pi\tilde{f})^2\} \\ sd[3n] &\leftarrow \{sd[3n]/(2\pi\tilde{f})^2\} \end{aligned} \quad (66)$$

where \leftarrow indicates replacement of the value of the left-hand term by the value of the right-hand term.

Algorithms ▼

5. Transform an analog filter to a digital filter via the bilinear transformation. The filters are transformed “in-place.” Transform the $sd[.]$ and $sn[.]$ coefficients for each section in a loop over the number of sections, N_s , (where n counts from 1 to N_s). For k , counting always from 1 to 3, define the temporary variable $c[.]$ in accordance with equation (67):

$$c[k] = sd[3(n-1) + k] \quad (67)$$

Form the temporary scaling variable r :

$$r = \sum_{k=1}^3 c[k] \quad (68)$$

Transform the coefficients for the n^{th} section:

$$\begin{aligned} sd[3(n-1) + 1] &= 1 \\ sd[3(n-1) + 2] &= \{2(c[3] - c[1])\}/r \\ sd[3(n-1) + 3] &= \{c[3] - c[2] + c[1]\}/r \end{aligned} \quad (69)$$

Apply precisely the same transformations to the $sn[.]$ as were delineated in Equations (67) through (69), above, for the $sd[.]$.

Application

The IIR filters designed by the package addressed in this chapter are generally multi sectioned. Each section consists of a set of three real-valued “numerator” coefficients, $sn[k]$, and two real-valued “denominator” coefficients, $sd[k]$. For implementation simplicity, although $sd[k]$ contains only two coefficients, space for three is allocated, and it is indexed as if three coefficients were used (the phantom coefficient in $sd[k]$ is set to 0). Thus, the coefficients for section 1 are found in $sn[1]$ through $sn[3]$ and $sd[1]$ through $sd[3]$, those for section 2 are found in $sn[4]$ through $sn[6]$ and $sd[4]$ through $sd[6]$, and so forth for a total of N_s sections. $sd[1]$, $sd[4]$, $sd[7]$, ... are phantom. (The IIR design process takes the user-specified type of filter and its parameters and determines the correct number of sections, N_s , and these sets of coefficients for each section.) The algorithm for applying an IIR filter to the time series, $x[n]$, of waveform data is recursive in that

outputs from the application of one filter section serve as inputs to the next. Arrays $a_k[n]$ and $b_k[n]$ serve as temporary local memories to convey intermediate filtered information from one section to the next and from one time-domain sample to the next. The overall filter application processing consists of a loop over the input waveform data samples, $x[i]$, and replaces these sample by sample with the filtered output. At the outset (that is, for input sample $x[i=1]$) the filter state memories, $a_k[n]$ and $b_k[n]$ are set to 0:

$$a_1[1] = \dots = a_1[N_s] = a_2[1] = \dots = a_2[N_s] = 0 \quad (70)$$

$$b_1[1] = \dots = b_1[N_s] = b_2[1] = \dots = b_2[N_s] = 0 \quad (71)$$

where the number of sections is N_s . An integer valued index, m , is set to 1; the value of the variable *input* is set to the first input datum, as in: $input = x[1]$, and the following equations are applied, in the order presented, in a loop over the number of sections, N_s :

$$output = input \cdot sn[m] + \sum_{k=1}^2 (a_k[n] \cdot sn[m+k]) - \sum_{k=1}^2 (b_k[n] \cdot sd[m+k]) \quad (72)$$

$$a_2[n] = a_1[n] \quad (73)$$

$$b_2[n] = b_1[n] \quad (74)$$

$$a_1[n] = input \quad (75)$$

$$b_1[n] = output \quad (76)$$

$$m \leftarrow m + 3 \quad (77)$$

(m is additively increased by three to access the next section's coefficients; \leftarrow indicates replacement of the value of the left-hand term by the value of the right-hand term)

$$input = output \quad (78)$$

Algorithms ▼

that is, for $n = 1, \dots, N_s$, inclusive, Equations (72) through (78) are applied, then equations (72) through (78) are again applied, and so forth for N_s passes at which time the original input datum $x[1]$ is set equal to the value currently held by the variable *output*. The next datum, $x[2]$ is processed by resetting m to unity, setting the value of the variable *input* equal to $x[2]$, and then processing equations (72) through 86 for N_s applications in sequence after which the original input datum $x[2]$ is set equal to the value currently held by the variable *output*. In processing sample $x[2]$ the filter state memories, $a_k[n]$ and $b_k[n]$ are generally no longer zero-valued. This process is continued until all of the input sample, $x[1]$ through $x[I]$, has been replaced by filtered values.

In general the phase response of an IIR filter is strongly dependent on frequency, and consequently, a single pass application of a multi-section IIR filter is likely to distort the waveform's shape. Those spectral components passed by the filter will have time delays, which are frequency dependent, and consequently, the superposition of these components will be out of time alignment causing the waveform's shape to be distorted. (The time alignment properties of the spectral components rejected by the filter are not important.) Where waveform shape is important, the user can request that the IIR filter be "zero-phase." When this request is made (by setting the filter specification parameter *zp* to unity) the IIR filter is applied to the data twice; first the input waveform data $x[.]$ are all filtered in accordance with the description immediately above, then the filtered data $x[.]$ are again filtered with the same IIR filter. However, now the time-series $x[i]$ is accessed in the reverse order, where i counts down from I down to 1. All other aspects of the filter's application are identical to its forward application for $i = 1, \dots, I$. The output from this second application, $x[.]$, is largely free of the waveform distortions. The application of a "zero-phase" IIR filter requires twice as much processing as the normal IIR filter of the same type, number of poles, and cutoff frequency(ies); however, the "zero-phase" IIR filter also provides twice the attenuation of the normal IIR filter.

Finite Impulse Response Filters

A finite impulse response (FIR) filter is a set of weights, and a time series can be filtered by convolving the weights and data. FIR filters can achieve a perfectly flat phase versus frequency response in the bandpass and, if allowed to be sufficiently long in the time domain, can achieve remarkably rapid transition from the bandpass to the stop band. However, the FIR filters are acausal, that is, they combine “future” and “past” information about the waveform to estimate each filtered sample.

FIR bandpass filtering is implemented in the frequency domain for efficiency. The input time domain waveform data, $x(n)$, are Fourier transformed to the frequency-domain, $X(k)$. The frequency-domain data values are conjugate-cross-multiplied by the frequency domain FIR bandpass filter response function $H(k)$, and then this complex-valued cross-product, $X(k)H^*(k)$, is inverse Fourier transformed back to the time domain to obtain the properly filtered time domain, $y(n)$, samples:

$$y(M + n_0 + n) = \sum_{k=0}^{K-1} X(k; n_0) \cdot H^*(k) \cdot e^{2\pi\sqrt{-1}(k \cdot n)/K} \quad (79)$$

K must be greater than the filter length, $2M + 1$ and must also be a power of 2 ($K = 2^P$; P a positive integer). For the filtered time series $y(n_0 + n)$ only the $(K - 2M)$ samples with indices $n = n_0$ through $n = (n_0 + K - 2M - 1)$ are valid. n_0 is the initial index for each time segment of filtering. The Fourier transformations into the frequency domain are given by equations (80) and (81):

$$H(k) = \frac{1}{K} \cdot \sum_{n=0}^{K-1} h(n) \cdot e^{-2\pi\sqrt{-1}(k \cdot n)/K} \quad (80)$$

$$X(k; n_0) = \frac{1}{K} \cdot \sum_{n=0}^{K-1} x(n_0 + n) \cdot e^{-2\pi\sqrt{-1}(k \cdot n)/K} \quad (81)$$

Because K must be greater than the filter length, $(K - 2M - 1)$ zeros are appended to the time domain filter, h , before being transformed into the frequency domain. $H(k)$ is generated only once for each filter and saved.

Algorithms ▼

The number of filtered data samples that are produced using this process is $K - 2M$. Overlapping segments of input waveform data, $x(n)$, are processed to obtain the desired filtered output, $y(n)$. This is accomplished by advancing the initial index, n_0 , in the algorithm as follows:

$$n_0 = 0, (K - 2M), 2*(K - 2M), 3*(K - 2M), \dots$$

Each filter is scaled in such fashion that an input process consisting of a noise-free constant frequency unit amplitude sine wave in the bandpass of the filter would lead to a filtered sinusoidal waveform time-series, which is also of unit amplitude. Thus, the FIR bandpass filters are effectively unity gain. The filters have a maximum stop-band response at least 50 dB below the in-band response with a transition region from bandpass to stop-band slightly less than $\Delta f = \text{sampling rate} * (0.003)$.

FIR filters are implemented in the *DFX* application (see [\[Wah96a\]](#) and [\[Wah96b\]](#)).

INTERPOLATION

Closest Sample Interpolation

Closest sample interpolation is used to estimate the value of a function, $x(t)$, between the closest two known values of the function, $x(t_1)$ and $x(t_2)$, where $t_1 \leq t < t_2$. Using two samples, the function is assumed to be equal in value at time t to the value of the function at either t_1 or t_2 depending upon which time difference is smaller ($t - t_1$) or $(t_2 - t)$, respectively.

Closest sample interpolation is implemented in the *DFX* application (see [\[Wah96a\]](#) and [\[Wah96b\]](#)).

Linear Interpolation

Linear interpolation is used to estimate the value of a function, $x(t)$, between the closest two known values of the function, $x(t_1)$ and $x(t_2)$, where $t_1 \leq t < t_2$. Using two samples, the function is assumed to be linear between the values provided at t_1 and t_2 ; the interpolated value is assumed to lie on the line joining $x(t_1)$ and $x(t_2)$.

Linear interpolation is used to estimate the values of missing or erroneous data points during data quality checks.

The equation for linear interpolation is as follows:

$$x(t) = x(t_1) + (t - t_1) \cdot \frac{x(t_2) - x(t_1)}{t_2 - t_1} \quad (82)$$

Linear interpolation is implemented in the *DFX* application (see [\[Wah96a\]](#) and [\[Wah96b\]](#)).

Cubic Interpolation

Cubic interpolation is used for estimating the value of a time series at time points that are not coincident with the times of the actual available samples. The lowest order process that can yield an interpolated estimate whose magnitude exceeds the absolute values of the four “nearest-neighbor” existing samples is used to develop the estimate. A cubic interpolation also yields smaller phase jitter than a linear interpolation.

Cubic interpolation is applied in resampling waveforms for delay and sum beam-forming, for resampling cross-correlation functions for realignment purposes, and for generating F-statistics.

The sample estimate for $x(t)$ using cubic interpolation is as follows:

$$x(t) = a_0 + [a_1 \varepsilon] + [a_2 \varepsilon^2] + [a_3 \varepsilon^3] \quad (83)$$

where $n\delta t \leq t < (n+1)\delta t$,

Algorithms ▼

δt is the sample interval (that is, inverse of the sampling rate),

$$\varepsilon = \frac{t - n\delta t}{\delta t}, \quad 0 \leq \varepsilon < 1, \quad (84)$$

$$a_0 = x(n), \quad (85)$$

$$a_1 = \frac{-2x(n-1) - 3x(n) + 6x(n+1) - x(n+2)}{6}, \quad (86)$$

$$a_2 = \frac{x(n-1) - 2x(n) + x(n+1)}{2}, \text{ and} \quad (87)$$

$$a_3 = \frac{-x(n-1) + 3x(n) - 3x(n+1) + x(n+2)}{6}. \quad (88)$$

This equation gives the originally sampled values $x(n)$ whenever t lies on a time point, which is exactly an original sample. Where t lies between samples n and $n+1$, $0 < \varepsilon < 1$, the estimated value of x at t , $x(t)$, can assume values larger in magnitude than the magnitudes of any of the four samples used in making the estimate.

Cubic interpolation is implemented in the *DFX* application (see [\[Wah96a\]](#) and [\[Wah96b\]](#)).

Two-dimensional Interpolation

Two-dimensional interpolation is used for estimating the value of a two-dimensional function at coordinates that are not coincident with the grid of the available samples.

Two-dimensional interpolation is accomplished using a bi-cubic spline algorithm that allows interpolated values larger (or smaller) than the surrounding values. The data must be regularly spaced in both rectangular coordinates, with identical resolution in both coordinates. A typical application is to estimate the actual position (\bar{s}_x, \bar{s}_y) of the peak value of an f - k spectrum where the function $F(s_x, s_y)$ has been calculated at gridpoints $F(i, j)$; where $s_x = i * \Delta s$, $s_y = j * \Delta s$; Δs is the resolution in each of the rectangular slowness coordinates; and i and j are integers, which run

over the interval $-l, -l+1, \dots, -1, 0, 1, \dots, l-1, l$. In general the actual peak position, (\bar{s}_x, \bar{s}_y) , will not lie on a gridpoint (\bar{i}, \bar{j}) ; rather it will be at some intermediate position in each coordinate. On the assumption that the resolution Δs is sufficiently small that the actual peak position is within a resolution unit of the gridpoint position (\bar{i}, \bar{j}) for which $F(s_x, s_y)$ is largest, the following algorithm will yield an estimate of the incremental change $(\delta x, \delta y)$ from the slowness at the gridpoint to the slowness at the actual position of the peak (\bar{s}_x, \bar{s}_y) .

Consider the three-by-three subregion of $F(s_x, s_y)$ centered on the gridpoint position for which F is largest:

$$\begin{aligned} &F[(\bar{i}-1)*\Delta s, (\bar{j}-1)*\Delta s] \\ &F[(\bar{i}-1)*\Delta s, (\bar{j})*\Delta s] \\ &F[(\bar{i}-1)*\Delta s, (\bar{j}+1)*\Delta s] \\ &F[(\bar{i})*\Delta s, (\bar{j}-1)*\Delta s] \\ &F[(\bar{i})*\Delta s, (\bar{j})*\Delta s] \\ &F[(\bar{i})*\Delta s, (\bar{j}+1)*\Delta s] \\ &F[(\bar{i}+1)*\Delta s, (\bar{j}-1)*\Delta s] \\ &F[(\bar{i}+1)*\Delta s, (\bar{j})*\Delta s] \\ &F[(\bar{i}+1)*\Delta s, (\bar{j}+1)*\Delta s] \end{aligned}$$

For convenience of notation, consider these nine samples of the function F as follows: $\Phi(-1,-1)$, $\Phi(-1,0)$, $\Phi(-1,1)$, $\Phi(0,-1)$, $\Phi(0,0)$, $\Phi(0,1)$, $\Phi(1,-1)$, $\Phi(1,0)$, and $\Phi(1,1)$, respectively. In terms of Φ , five derived parameters, a , b , c , d , and e , can be defined as follows:

$$a = \left\{ \frac{\Phi(-1,-1) + \Phi(-1,0) + \Phi(-1,1)}{6} \right\} - \left\{ \frac{\Phi(0,-1) + \Phi(0,0) + \Phi(0,1)}{3} \right\} + \left\{ \frac{\Phi(1,-1) + \Phi(1,0) + \Phi(1,1)}{6} \right\} \quad (89)$$

$$b = \left\{ \frac{\Phi(-1,-1) - \Phi(-1,1)}{4} \right\} - \left\{ \frac{\Phi(1,-1) - \Phi(1,1)}{4} \right\} \quad (90)$$

$$c = \left\{ \frac{\Phi(-1,-1) + \Phi(0,-1) + \Phi(1,-1)}{6} \right\} - \left\{ \frac{\Phi(-1,0) + \Phi(0,0) + \Phi(1,0)}{3} \right\} + \left\{ \frac{\Phi(-1,1) + \Phi(0,1) + \Phi(1,1)}{6} \right\} \quad (91)$$

Algorithms ▼

$$d = - \left\{ \frac{\Phi(-1, -1) + \Phi(-1, 0) + \Phi(-1, 1)}{6} \right\} + \left\{ \frac{\Phi(1, -1) + \Phi(1, 0) + \Phi(1, 1)}{6} \right\} \quad (92)$$

$$e = - \left\{ \frac{\Phi(-1, -1) + \Phi(0, -1) + \Phi(1, -1)}{6} \right\} + \left\{ \frac{\Phi(-1, 1) + \Phi(0, 1) + \Phi(1, 1)}{6} \right\} \quad (93)$$

The terms a and c are measures of the degree to which the peak's actual position is not centered along the s_x (i) and s_y (j) directions, respectively. Similarly, the terms d and e are measures of the degree to which the peak's actual position is not centered along the s_x (i) and s_y (j) directions, respectively. The b term is a measure of "twist" of the peak's shape with respect to the s_x and s_y directions. That is, if the peak were cut by a plane parallel to the (s_x, s_y) plane at approximately half peak height, would the closed figure in this "cutting plane" exhibit an angle with respect to the s_x and s_y directions, or would the figure, for example, an ellipsoid appear to have its major and minor axes in alignment with the s_x and s_y directions? In terms of a , b , and c the quantity D can be defined as

$$D = b^2 - 4 \cdot a \cdot c \quad (94)$$

and where D is not too small in absolute value (that is, $> 10^{-10}$) the incremental adjustment, $(\delta x, \delta y)$, to the gridpoint peak position can be estimated as follows:

$$\delta x = \left\{ \frac{(2 \cdot a \cdot e - b \cdot d)}{D} \right\} \cdot \Delta s \quad (95)$$

$$\delta y = \left\{ \frac{(2 \cdot c \cdot d - b \cdot e)}{D} \right\} \cdot \Delta s \quad (96)$$

and the final estimate of actual peak position, (\bar{s}_x, \bar{s}_y) , is given by

$$\bar{s}_x = (i \cdot \Delta s) + \delta x \quad (97)$$

$$\bar{s}_y = (j \cdot \Delta s) + \delta y \quad (98)$$

In case the absolute value of the parameter D is less than 10^{-10} , the incremental adjustment, $(\delta x, \delta y)$, is set to $(0, 0)$, and the “actual” peak position is estimated as the position of the gridpoint of greatest F value, namely the coordinates of the gridpoint, (\bar{i}, \bar{j}) :

$$\bar{s}_x = (\bar{i} \cdot \Delta s) \quad (99)$$

$$\bar{s}_y = (\bar{j} \cdot \Delta s) \quad (100)$$

Two-dimensional interpolation is implemented in the *DFX* application (see [\[Wah96a\]](#) and [\[Wah96b\]](#)).

BEAM GENERATION

A beam is a sum, or stack, of data channels from an array station in which the data channels are time shifted to enhance coherent energy with a particular slowness and azimuth. Lags are calculated for each raw data channel based on the position of the sensor relative to the reference location (**site.dnorth** and **site.deast**) and the phase velocity vector appropriate to the beam being created:

$$\text{lag}_i = (\text{dnorth}_i \cdot s_n + \text{deast}_i \cdot s_e) \cdot \text{samprate} \quad (101)$$

where dnorth_i and deast_i are the north and east offsets in kilometers of element i relative to the array reference point (stored in the **site** database table), s_n and s_e are the north and east components of slowness in seconds per kilometer, and *samprate* is the sampling rate of the waveform data in samples per second (stored in the **wfidsc** database table). When forming beams from seismic data, the lag_i are rounded to integers. That is, closest sample interpolation is applied. For infrasonic data, however, the generally non-integral valued lag_i are used, and the waveform data are interpolated (see [“Interpolation” on page 186](#)).

Caution must be used when forming beams from masked data. Every beam has an associated “beam norm” that is a stack of the masks of the waveforms used to create the beam. This stack is used to normalize the beam on a point-by-point

Algorithms ▼

basis because the number of elements contributing to the beam changes as a function of time. For regions where all sensor-channels are masked, the beam value is set to zero.

Three types of beam can be formed; coherent, incoherent, and rms. Coherent beams are formed by delaying the signal from each of the J elements in an array of sensors, x_i , by an amount corresponding to the phase velocity and azimuth of the desired beam, summing the delayed traces, normalizing, and then filtering:

$$\text{beam}(n)_{\text{coherent}} = \text{filt} \left\{ \frac{\sum_{i=1}^J x_i(n - \text{lag}_i)}{\sum_{i=1}^J \text{norm}_i(n - \text{lag}_i)} \right\} \quad (102)$$

Because coherent beamforming is a linear process, the filtering and summation over sensor-channels can be interchanged, thereby reducing the processing burden of the filtering by a factor of $(J-1)$.

Incoherent beams are formed by filtering each array element, delaying the elements as described for coherent beams, rectifying, summing, and normalizing. Because the elements are rectified before summing—a nonlinear process—each element must be filtered separately, which causes incoherent beams to be much more computationally expensive to create than coherent beams. The algorithm for forming incoherent beams is as follows:

$$\text{beam}(n)_{\text{incoherent}} = \frac{\sum_{i=1}^J |\text{filt}\{x_i(n - \text{lag}_i)\}|}{\sum_{i=1}^J \text{norm}_i(n - \text{lag}_i)} \quad (103)$$

The rms beam is similar to an incoherent beam. Each element is filtered, and delayed, but the data values are squared rather than rectified before summing, and the square root of the normalized values are summed:

$$\text{beam}(n)_{\text{rms}} = \frac{\sum_{i=1}^J \sqrt{(\text{filt}\{x_i(n - \text{lag}_i)\})^2}}{\sum_{i=1}^J \text{norm}_i(n - \text{lag}_i)} \quad (104)$$

Beams are calculated in the *DFX* application (see [\[Wah96a\]](#) and [\[Wah96b\]](#)).

NORMALIZED CROSS-CORRELATION FUNCTIONS

A normalized cross-correlation function of two time series is an estimate of the coherence between the two time series over a number of lag values. The two time series must have the same sample rate and are time aligned at zero lag.

High values of coherence between the data from two array elements indicates that nonrandom energy is present.

Normalized cross-correlation functions between pairs of data channels are generated using a circular-convolution approach based upon the Fast Fourier Transform (FFT). The results of the calculation are identical (to better than 1 part in 10^5) to the results that would have been obtained via the time-domain application of the normalized cross-correlation.

The N time-domain values of the data segment for channel i , $x_i(n)$, beginning at the n_0 th sample are transformed to the frequency-domain by the following equation:

$$X_i(k) = \frac{1}{K} \cdot \sum_{n=0}^{K-1} x'_i(n + n_0) \cdot e^{-2\pi\sqrt{-1}(n \cdot k)/K} \quad (105)$$

where $x'_i(n) = x_i(n)$ for $n_0 \leq n \leq n_0 + (N - 1)$, and $x'_i(n) = 0$ otherwise. K is the total length of the FFT (defined below), and the index of the frequency-domain values is k .

Algorithms ▼

Convolution requires that the data segments be shifted in time (lagged) relative to one another, so the second data segment must have enough data values to accommodate the largest lag, L , desired. L is computed using the maximum slowness, S_{\max} , the distance between sensors i and j (computed from `site.dnorth` and `site.deast` for each sensor), and the sample rate, `samprat`:

$$L = \text{integer}(2 + S_{\max} \cdot \Delta r \cdot \text{samprat} + 0.5) \quad (106)$$

where

$$\Delta r = \sqrt{(dnorth_i - dnorth_j)^2 + (deast_i - deast_j)^2} \quad (107)$$

and `integer()` is the largest integer of the term in the parentheses. The addition of 0.5 in (106) forces a rounding up and the addition of 2 ensures that enough lag values are available for interpolation. To simplify data processing and interpretation, the single largest value of Δr is used for all sensor pairs at a given array site, even in cases where the sensor elements, which are mostly separated at a given array site, are not to be used in the processing. (That is, no beam-group calls for that most widely spaced pair of sensors.)

The second data segment must have $N + 2L$ data values (L lag values on either end of the data segment). The $N + 2L$ time-domain values of the data segment for channel j , $x_j(n)$, beginning at the $(n_0 - L)$ th sample are transformed to the frequency-domain by

$$X_j(k) = \frac{1}{K} \cdot \sum_{n=0}^{K-1} x'_j(n + n_0 - L) \cdot e^{-2\pi\sqrt{-1}(n \cdot k)/K} \quad (108)$$

where $x'_j(n) = x_j(n)$ for $n_0 - L \leq n \leq n_0 + (N - 1) + L$, and $x'_j(n) = 0$ otherwise.

The value of K must be a power of 2 (for the FFT algorithm used) and must be large enough to include all of the data samples. Thus K must be selected such that K is a positive integer satisfying the conditions that $K \geq N + 2L$ and $K = 2^P$ where P is a positive integer.

The unnormalized cross-correlation function, $\tilde{\chi}_{ij}(\text{lag})$, is the inverse FFT of the conjugate inner product of the frequency-domain values of the data segments:

$$\tilde{\chi}_{ij}(\text{lag}) = \sum_{k=0}^{K-1} X_i^*(k) \cdot X_j(k) e^{2\pi\sqrt{-1}(\text{lag} \cdot k)/K} \quad (109)$$

where the lag is constrained to $-L \leq \text{lag} \leq L$.

To normalize the cross-correlation function, (109) is divided by the normalization terms P_i and $P_j(\text{lag})$:

$$\chi_{ij}(\text{lag}) = \frac{\tilde{\chi}_{ij}(\text{lag})}{P_i \cdot P_j(\text{lag})} \quad (110)$$

where

$$P_i = \sum_{n=n_0}^{n_0+(N-1)} |x_i(n)|^2 \quad (111)$$

$$P_j(\text{lag}) = \sum_{n=n_0}^{n_0+(N-1)} |x_j(n+\text{lag})|^2 \quad (112)$$

for channels i and j , respectively.

After a normalized cross-correlation function, (110), has been computed for each pair of sensors in an array station, the results can be combined in slowness space to identify the directions that the coherent energy is travelling as well as the phase velocity of this energy passage across the sensor array.

For each point in the slowness plane, $\Gamma(s_n, s_e)$, the matching lag for each normalized cross-correlation function can be calculated:

$$\text{lag}_{ij}(s_n, s_e) = ((dnorth_i - dnorth_j) \cdot s_n + (deast_i - deast_j) \cdot s_e) \cdot \text{samprate} \quad (113)$$

The slowness plane, $\Gamma(s_n, s_e)$, is populated by summing the contributions of each normalized cross-correlation function at each slowness coordinate:

$$\Gamma(s_n, s_e) = \frac{1}{\left[\frac{J \cdot (J-1)}{2} \right]} \cdot \sum_{i=1}^{J-1} \left\{ \sum_{j=i+1}^J \chi_{ij}(\text{lag}_{ij}(s_n, s_e)) \right\} \quad (114)$$

Algorithms ▼

where J is the number of sensor elements in the array station.

The lag_{ij} are generally not integer valued. Cubic interpolation is used to estimate the numerical value of the normalized cross-correlation coefficient, χ_{ij} , at the non-integer valued lag_{ij} given the sampled values of χ_{ij} at the discrete values of lag (see [“Cubic Interpolation” on page 187](#)).

Normalized cross-correlation functions are calculated in the *DFX* application (see [\[Wah96a\]](#) and [\[Wah96b\]](#)).

F-STATISTICS

F-statistics provide an alternative mechanism for estimating the spatial coherence among the bandpass filtered waveform segments from the sensors of an array station (in contrast to the arithmetic mean of the lag-aligned normalized cross-correlation functions, χ_{ij} , that estimate spatial coherence between the waveform segments for given pairs of sensors).

The F-statistic is a ratio of the time-averaged power in a coherent beam to the time-averaged power in the “residuals” summed over the same set of data channels used to form the beam. The “residuals” are the waveform data for each sensor from which the properly time-aligned normalized beam has been subtracted. If the original waveform data for a given sensor-channel were equal to the normalized beam, then the “residuals” for that channel would be zero. Thus, the “residuals” can be thought of as the incoherent noise amplitudes in the given sensor channel.

The F-statistic is defined by the following equation:

$$FS(s_n, s_e) = \left[\frac{J-1}{J} \right] \cdot \frac{\sum_{n=n_0}^{n_0+(N-1)} \left| \sum_{i=1}^J x_i(n + \text{lag}_i(s_n, s_e)) \right|^2}{\sum_{n=n_0}^{n_0+(N-1)} \sum_{i=1}^J \left| x_i(n + \text{lag}_i(s_n, s_e)) - \left[\frac{1}{J} \right] \cdot \sum_{j=1}^J x_j(n + \text{lag}_j(s_n, s_e)) \right|^2} \quad (115)$$

where the F-statistic is to be evaluated starting at the time sample index, n_0 , with N time-domain samples of waveform data in the calculation for each of J sensor-channels. The $\text{lag}_i(s_n, s_e)$ are the time lags required to steer the beam to the azimuth and slowness given by the slowness coordinates, (s_n, s_e) . The amplitude of the zero-mean bandpass filtered datum for the i^{th} sensor at the time sample index n is $x_i(n)$.

If each of the data segments used is detrended, offset to have zero mean, and scaled to have unit variance (i.e., total time-averaged power of unity), then (115) reduces to the following form:

$$FS(s_n, s_e) = \frac{(J-1) \cdot B(s_n, s_e)}{J^2 - B(s_n, s_e)} = (J-1) \cdot \frac{[B(s_n, s_e)/J^2]}{\{1 - [B(s_n, s_e)/J^2]\}} \quad (116)$$

where

$$B(s_n, s_e) = \sum_{n=n_0}^{n_0+(N-1)} \left| \sum_{i=1}^J x_i(n + \text{lag}_i(s_n, s_e)) \right|^2 \quad (117)$$

is the total time-averaged power in the beam.

When there is no noise, $B(s_n, s_e)$ approaches J^2 , and FS approaches infinity. In contrast, where no plane wave arrival is present and the data are random noise, $B(s_n, s_e)$ approaches J and FS approaches unity.

The numerical values of the lag terms, $\text{lag}_i(s_n, s_e)$, that align the data segments to the azimuth and slowness of the beam are given by

$$\text{lag}_i(s_n, s_e) = (dnorth_i \cdot s_n + deast_i \cdot s_e) \cdot \text{samprate} \quad (118)$$

where *samprate* is the sample rate in samples per second from the **wfdisc** database table, *dnorth_i* and *deast_i* are the positions of the i^{th} sensor relative to the array reference location from the **site** database table, and s_n and s_e are the north and east components of slowness for which the F-statistic is being evaluated. The $\text{lag}_i(s_n, s_e)$ are generally not integer valued. If the plane wave arrival at the sensor site is later in time than that plane waves arrival at the reference site, they are pos-

Algorithms ▼

itive valued. Where the $\text{lag}_i(s_n, s_e)$ value is not an integer, cubic interpolation is used to produce a resampled time series that is correctly aligned in time for the given lags (see [“Cubic Interpolation” on page 187](#))

F-statistics are calculated in the *DFX* application (see [\[Wah96a\]](#) and [\[Wah96b\]](#)).

STA/LTA DETECTORS

STA/LTA detectors are ratios of averages; the STA (short-term-average) is a time-average of some function of the data (for example, amplitude, power, or energy) over a short time period relative to the LTA (long-term-average); and the LTA is an average of the same function over a long time period relative to the STA. When the input process is simple random noise, the ratio fluctuates about unity; however, when a transient process with excess energy is encountered, the “energy” contained in the STA window rapidly grows larger compared with the “energy” contained in the time-trailing and much longer LTA. By comparing the ratio, sample by sample, with a fixed threshold value, transients are detected.

STA/LTA detectors are implemented in the *DFX* application (see [\[Wah96a\]](#) and [\[Wah96b\]](#)). Several options exist for applying STA/LTA detectors using the *DFX* application. The STA/LTA detectors used for detecting signals in seismic and hydroacoustic data use a running STA, and the LTA is computed recursively using previously computed STA values. By contrast, the STA/LTA detectors used for detecting signals in infrasonic data use running short-term and generally time-lagged long-term averages, which are independently calculated.

STA/LTA detectors are implemented in the *DFX* application (see [\[Wah96a\]](#) and [\[Wah96b\]](#)).

Seismic and Hydroacoustic STA/LTA Detectors

The STA/LTA detector for seismic and hydroacoustic data uses a running average for the STA and a recursive average for the LTA. The LTA “lags” the STA in time by half the width of the STA window.

In the following description, an algebraic detail of the implementation has been omitted for clarity. This detail addresses the issue that for seismic waveform data, which have been beamformed, the number of sensor-channels, J , contributing to each beam sample, $x(n)$, may vary in accordance with the quality of the waveform data in each sensor-channel. This sample-by-sample variation in J , namely, $J(n)$, must be accounted for in properly normalizing each sample of $x(n)$. Also, for both seismic and hydroacoustic waveform data, the number of good waveform samples of $x(n)$ available in any time interval of fixed length corresponding to M samples may be less than M samples. This issue must also be accounted for in the STA/LTA detector algorithm. Both issues tend to complicate the algebra considerably. A full explanation requires an examination of detail at the level of pseudocode.

For a time series $x(n)$, where n is the sample index and $x(n)$ is the amplitude at sample n , the initial value of the STA, $stav$, is calculated as

$$stav(S/2) = \frac{1}{S} \cdot \sum_{s=0}^{S-1} |x(s)| \quad (119)$$

where S is the number of samples in the STA window.

A recursion equation is used to compute subsequent values of the STA:

$$stav(k) = stav(k-1) + \frac{1}{S} \cdot \left[\left| x\left(k + \frac{S}{2}\right) \right| - \left| x\left(k-1 - \frac{S}{2}\right) \right| \right] \quad (120)$$

where $\left(\frac{S}{2}\right) \leq k \leq (N-1) - \left(\frac{S}{2}\right)$, and N is the number of available samples in the time series. Samples that have been “masked” in the quality check are not included in the averages. For the end-segment intervals, $k < [S/2]$ and $k > (N-1) - [S/2]$, the following assignments are made:

$$stav(k) = stav\left(\frac{S}{2}\right) \quad (121)$$

and

$$stav(k) = stav\left((N-1) - \left(\frac{S}{2}\right)\right), \quad (122)$$

Algorithms ▼

respectively.

The LTA, $Itav(k)$ is computed recursively from the STA as

$$Itav(k) = \left(1 - \frac{1}{L}\right) \cdot Itav(k-1) + \left(\frac{1}{L}\right) \cdot stav(k-S) \quad (123)$$

where L is the effective number of samples in the LTA window. That is, the $Itav(k)$ is calculated as an "RC integration" of $stav$ estimates. RC integration is a one-pole form of infinite impulse response (IIR) low-pass filtering. This form can be physically built by taking the output signal from the terminals of a capacitor (C), which is supplied input signal through a series resistor (R). The capacitor is the "integrator" or "summer" and the series resistor provides the mechanism for isolating input from output and for setting the actual time constant, T , or the integrator ($T=RC$). In equation (123), the running index, k , of the $Itav$ is confined to the interval:

$$\left(\frac{S}{2}\right) \leq k \leq N - \left(\frac{S}{2}\right) \quad (124)$$

For the end-segment intervals, $k < [S/2]$ and $k > N - [S/2]$, the following assignments are made:

$$Itav(k) = Itav\left(\frac{S}{2}\right) \quad (125)$$

and

$$Itav(k) = Itav\left[N - \left(\frac{S}{2}\right)\right] \quad (126)$$

respectively.

The foregoing algebra depicted an L1 norm, namely the time averages of the waveform amplitude, which is the normal mode for seismic applications; the algorithm itself is coded to allow for either L1 or L2 norms, depending on the setting of a software parameter. In the case of the L2 norm, the time averages are of estimates of the waveform's squared amplitude, which is closely related to the wave-

form's instantaneous power. A data channel is considered to be in a detection state when the ratio of STA to LTA exceeds the specified threshold for that particular data channel.

Infrasonic STA/LTA Detectors

The STA/LTA detector used for infrasonic data differs from the STA/LTA detector used for seismic and hydroacoustic data: a time gap exists between the STA and the LTA windows. These special attributes are particularly useful where the transients sought are characterized by emergent onset. In addition the infrasonic STA/LTA detector can be set to any of four modes: L1 or L2 norm with either simple or complex architecture. The L1 and L2 norms have identical meaning to the seismic/hydroacoustic use.

The simple architecture maintains a fixed relationship between the time registration of the LTA window relative to the time registration of the STA window. In the complex architecture, the LTA window ceases time advance as soon as detection occurs, and the LTA window is only advanced to its normal time alignment (with respect to the STA window) after the STA window has advanced a sufficient time since the most recent detection. The default mode is L1 norm with simple architecture.

For a time series $x(m)$, where m is the sample index and $x(m)$ is the amplitude at sample m , the value of the STA, $stav(k)$, and the value of the LTA, $ltav(k)$, are given by

$$stav(k) = \frac{1}{S} \cdot \sum_{m=k}^{k+(S-1)} \left| x\left(m - \left\lfloor \frac{S}{2} \right\rfloor\right) \right| \quad (127)$$

$$ltav(k) = \frac{1}{L} \cdot \sum_{m=k}^{k+(L-1)} \left| x\left(m - \left(\left\lfloor \frac{S}{2} \right\rfloor + G + L\right)\right) \right| \quad (128)$$

Algorithms ▼

where S is the number of samples in the STA window, L is the number of samples in the LTA window, G is the number of samples by which the most recent time sample in the LTA window lags the most ancient sample in the STA window, and N is the number of available waveform data samples $x(n)$. The logical indices of both the stav and the ltav run over the following interval:

$$\left(\left\lceil\frac{S}{2}\right\rceil + G + L\right) \leq k \leq (N - 1) - \left\lceil\frac{S}{2}\right\rceil \quad (129)$$

That is, the samples of the stav and ltav arrays with the same logical indices are the samples to be compared for detection purposes. In total $[(N+1)-(S+G+L)]$ samples of each of the arrays stav and ltav are calculated. No values are calculated for either the stav or the ltav for logical indices:

$$k < \left(\left\lceil\frac{S}{2}\right\rceil + G + L\right) \quad (130)$$

or

$$k > (N - 1) - \left\lceil\frac{S}{2}\right\rceil \quad (131)$$

The actual index, k' , of the ltav and stav arrays spans the interval:

$$0 \leq k' \leq N - (S + G + L) \quad (132)$$

To interpret index k' relative to time, the initial offset of the logical index k to the actual index k' by $(L + G + [S/2])$ samples must be considered:

$$k' = k - \left(L + G + \left\lceil\frac{S}{2}\right\rceil\right) \quad (133)$$

Recursion equations (regardless of whether the actual index k' or logical index k is considered) are used to compute subsequent values of the STA and LTA:

$$\text{stav}(k + 1) = \text{stav}(k) + \left\langle \frac{1}{S} \cdot \left\{ \left| x\left(k + \left\lceil\frac{S}{2}\right\rceil\right) \right| - \left| x\left(k - \left\lceil\frac{S}{2}\right\rceil\right) \right| \right\} \right\rangle \quad (134)$$

$$Itav(k+1) = Itav(k) + \left\langle \frac{1}{L} \cdot \left\{ \left| x \left(k - \left\lceil \frac{S}{2} \right\rceil + G \right) \right|^2 - \left| x \left(k - \left\lceil \frac{S}{2} \right\rceil + G + L \right) \right|^2 \right\} \right\rangle \quad (135)$$

S is selected to be the average number of time-domain samples in a typical signal so that the normalized value calculated in the S window is generally free of “noise” energy. The “gap” G between the end of the L window and the beginning of the S window is typically set at a default value of $2S$ samples; however, for special applications the value of G can be set otherwise within the infra processing recipe. G is intended to keep the normalized value calculated in the L window free of “signal” energy in the STA window.

SEISMIC ARRIVAL-TIME ESTIMATION VIA AKAIKE INFORMATION CRITERION

The Akaike Information Criterion (AIC) method for estimating the arrival time of a transient signal arrival ([Kam92], [Kam94], [Kvæ95]) is discussed in [“Estimating Arrival Time” on page 22](#). The key feature of this method is that it seeks to estimate the arrival time by distinguishing the properties of a segment of waveform data presumed to be ambient noise from an adjacent segment of waveform data presumed to contain both signal and ambient noise. The signal’s arrival time is that point in time that joins the two waveform segments. Because the time of the signal’s arrival is not known previously, this parameter is sought over a suitably chosen time window. The test statistic, the AIC, has been formulated, which is monotonically decreasing with respect to the similarities in the properties of the two waveform segments. The arrival time is estimated to be that time for which the properties of the immediately proceeding waveform segment are maximally different from the properties of the immediately following waveform segment. The AIC method defines both the properties and the test statistic.

Several issues should be considered when implementing the AIC method:

- choosing the AIC prediction model
- constructing the noise and signal models
- filtering

Algorithms ▼

- evaluating the models

These issues are discussed in detail in the following sections:

Choosing AIC Prediction Models

When selecting Models 1 and 2, decide first which processes to model, and second, to which parts of the Onset Search Window they should be applied. If the processes adhere to the assumptions behind the AIC the more powerful approach will be to employ models of both; then the determination of onset is based on how much the time series is like or unlike each of two processes. In this context, the noise model is generally used for Model 1 and the signal model is used for Model 2. This choice is expected to provide the best over-all prediction of the data in the Onset Search Window (hence the minimum value of the AIC) by two models to be achieved when the segment before the onset is predicted by the noise model and the segment after the onset by the signal model. The opposite selection is just as clearly counterproductive in decreasing the AIC toward the onset time and is not allowed in this implementation.

However, as previously discussed, the signal generally adheres less well than the noise to the statistical assumptions of the AIC. Especially in low snr situations, a decent model of the signal may be difficult or impossible to create. Use of a poor model of a poorly behaved signal process in the AIC may only dilute the discriminatory virtues of a good model of a much better behaved noise process. Thus another popular formulation of the AIC is using the noise model for both Models 1 and 2. Similar logic can be applied to higher snr and statistically better behaved signals to suggest the use of the signal model for both Models 1 and 2, although this practice is not common.

In this implementation the combination for Models 1 and 2 is specified through a parameter of the onset recipe. All combinations except signal/noise may be specified for Model 1/Model 2. The default combination is noise/noise.

Constructing Noise and Signal Models

In principle, the noise can be modeled from any segment of signal-free data. However, given the limited sense in which it may be considered stationary, part of the skill in constructing a noise model is isolating a waveform segment that is short enough and near enough to the onset that its statistics are representative of the noise around the onset, yet long enough to be statistically meaningful. For maximum effectiveness, the segment must not contain any signal so it is best selected in advance of the seed detection. In contrast, the signal is temporally localized so choosing a time interval from which to construct the signal model is more limited. Accurate determination of the signal segment is difficult; a more practical goal is to identify a window containing most or all of the signal, even if it contains signal-free sections. The signal model is likely to be a crude predictor of the signal, but as long as the model predicts the signal significantly better than the noise model and predicts the noise significantly worse than the noise model, its inclusion can enhance the ability of the AIC to identify the onset time.

In this implementation, the AR models are computed from subsegments of the Onset Search Window. The lengths of these segments and their positions relative to the seed detection are controlled by onset-recipe parameters. The models themselves are damped least-squares AR models; the onset-recipe contains parameters controlling details of the model and the modeling procedure.

Filtering

The AIC method is entirely premised on the noise-only and signal regions being statistically distinguishable. Thus it can work effectively on an unfiltered series. However, the strength of the method diminishes as the statistical properties of the noise and signal regions converge. This can happen either because the signal and noise processes themselves are statistically similar—which cannot be cured—or because the signal is weak compared to the noise. This latter situation can be improved dramatically by applying filters that preferentially pass only the higher snr frequency band of the signal.

Algorithms ▼

This implementation provides the option to specify and apply a suite of filters to the onset trace before the prediction models are generated and the AIC is computed. The apparent snr of each filtered trace is calculated by using short-term and long-term averages, and the highest snr trace is selected. The nature of the snr measure and details of its calculation are specified in a snr-filter recipe. The filters themselves are specified in filter recipes that may be stored in the recipe table format.

A host of caveats can be observed when filtering in the context of onset time estimation. Acausal filters, such as finite impulse response filters, will spread signal energy backward in time to the degree to which their stop bands overlap with the energetic band of the signal; thus if the cut-off frequencies are poorly chosen early arrival times estimations may result. In contrast, causal filters such as one-pass infinite impulse filters have non-linear phase, particularly outside of the pass-band, that tends to delay higher frequency energy near the pass-band; this can severely alter the shape of the phase wavelet and can lead to late estimates of arrival time. Filter parameters must be carefully selected to avoid these effects.

Evaluating Models

The conventional wisdom is that Model 1 be evaluated in the forward time direction and Model 2 in the reverse time direction to avoid edge effects when evaluating near the pivot point k . In this implementation Model 1 is always evaluated in the forward direction; for the default configuration Model 2 is evaluated in the reverse direction, but this may be specified explicitly.

FREQUENCY-WAVENUMBER ANALYSIS

Frequency-wavenumber (f - k) analysis is used to estimate the azimuth, slowness, and their errors of signals detected in array data. The array data are transformed into a frequency-wavenumber power spectrum, the peaks of which are located at the azimuths and slownesses of spatially coherent waveform data.

The wide-band f-k algorithm of [Kvæ86] is used as a basis for computation. The f-k power spectrum $P(s_n, s_e)$ is computed from equation (136).

$$P(s_n, s_e) = \frac{\sum_{f=f_1}^{f_2} \left| \sum_{i=1}^J F_i(f) \cdot e^{2\pi\sqrt{-1}f(s_n \cdot dnorth + s_e \cdot deast)} \right|^2}{J \cdot \sum_{f=f_1}^{f_2} \left\{ \sum_{i=1}^J F_i(f)^2 \right\}} \quad (136)$$

where

- $X_i(f)$ is the Fourier transform of $x_i(n)$ —the waveform data for the i^{th} sensor array element at time sample n .
- $F_i(f) = \sqrt{X_i^*(f) \cdot X_i(f)}$ is the Fourier amplitude of the i^{th} sensor array element at frequency f ; s_e and s_n are the east-west and north-south components of the vector slowness, respectively, for which P is being calculated.
- $deast_i$ and $dnorth_i$ are the east-west and north-south coordinates, respectively, of the of the i^{th} sensor array element relative to the reference station (found in the **site** database table).
- f_1 and f_2 are the low and high frequency limits, $fklof$ and $fkhi$, respectively, set in the f-k spectrum processing recipe;
- and J is the number of sensor elements in the array.

In general, the waveform data, $x_i(n)$, are first bandpass filtered to the limits flo and fhi , also set in the f-k spectrum processing recipe. The calculation of P is sensible only if the following conditions are satisfied:

$$fklof \leq fkhi \quad (137)$$

$$flo \leq fklof \leq fhi \quad (138)$$

$$flo \leq fkhi \leq fhi \quad (139)$$

Algorithms ▼

The time duration and time centering of the f-k spectrum are governed by the parameters t_lead , t_lag , and $t_arrival$. That is, the Fourier spectrum, $X_i(f)$, is calculated from waveform data for the i^{th} sensor-channel for the time block from time t_1 until time t_2 , where $t_1 = (t_arrival - t_lead)$ and $t_2 = (t_arrival + t_lag)$. The overall time duration, T , of the each Fourier spectrum, $X_i(f)$, is given by $T = (t_2 - t_1) = (t_lead + t_lag)$, and this time duration governs the frequency resolution, $\Delta f (= 1/T)$, of each spectrum. This frequency resolution defines the picket indexing of the outer sum over “frequency” in equation (136), above. That is, the sum extends from integer-valued indices k_1 to k_2 , where $k_1 = [(f_1/\Delta f) + 0.5]$ and $k_2 = [(f_2/\Delta f) + 0.5]$, and $[x]$ represents the greatest integer in x .

The slowness coordinates of the peaks in the f-k power spectrum are refined through two-dimensional interpolation (see [“Two-dimensional Interpolation” on page 188](#)). The refined coordinates are then used to calculate the azimuth and slowness of the signal’s spatially coherent plane wave energy. The azimuth is calculated from the arc-tangent of the slowness coordinates of the peaks (Se_p , Sn_p), and the slowness is obtained from the amplitude of the slowness vector between the slowness origin at (0, 0) and the peak:

$$azimuth = \begin{cases} \text{atan}\left(\left|\frac{Se_p}{Sn_p}\right|\right) & Se_p, Sn_p \geq 0 \\ 90 + \text{atan}\left(\left|\frac{Se_p}{Sn_p}\right|\right) & Se_p \geq 0, Sn_p < 0 \\ 180 + \text{atan}\left(\left|\frac{Se_p}{Sn_p}\right|\right) & Se_p < 0, Sn_p < 0 \\ 270 + \text{atan}\left(\left|\frac{Se_p}{Sn_p}\right|\right) & Se_p < 0, Sn_p \geq 0 \end{cases} \quad (140)$$

$$slow = \sqrt{Se_p^2 + Sn_p^2}$$

Azimuth and *slow* are stored in the **arrival** database table in degrees with respect to North and in seconds per degree of distance on the earth’s surface, and they are stored in the **detection** database table as *seaz* (also in degrees with respect to North) and *slow* in seconds per kilometer. The **detection** database table entries are never revised after they are initially written by the automatic processing; in contrast those in the **arrival** database table may be revised during interactive analysis.

F-k analysis is performed within the *DFX* application (see [\[Wah96a\]](#) and [\[Wah96b\]](#)).

Error Estimation

Theoretical error estimates for slowness and azimuth are based on the vector slowness resolution of the spectrum for the frequency band in which the processing is performed (*fklof* to *fkhi*) and on the quality of the measurement as reflected in the *fstat* value (which is itself a measure of the degree of spatial coherence). To estimate the resolution in vector slowness, the theoretical beam pattern is analyzed to find the difference between the azimuth and magnitude slowness values at the peak power and their values at 1 dB below the peak. For a symmetric array, the 1 dB contour is a circle. If the radius of the circle is dk , the errors in azimuth, da , and velocity, dv , at the center frequency, f_c , are as follows:

$$\delta a = \text{asin}\left(\frac{\delta k}{f_c}\right) \quad (141)$$

$$\delta v = \frac{v^2}{f_c \cdot \delta k} \quad (142)$$

An f-k quality measure (*fkqual*) is calculated to account for additional errors caused by noise, interfering signals, and deviation of the signal from the assumed plane-wave. When the amplitude of the second highest peak is more than 6 dB less than that of the highest peak in the f-k plane, *fkqual* is set to 1. *fkqual* set to 2, 3, and 4 when the difference between the highest and next highest peaks is 4–6 dB, 2–4 dB, and 0–2 dB, respectively.

The vector slowness resolution error and f-k quality are combined to obtain estimates of the slowness and azimuth errors. When *fkqual* is 1, the resolution error is interpreted to be 2 standard deviations (2 s.d.). When *fkqual* is 2, the error is 1.5 s.d., and when *fkqual* is 3, the resolution error is interpreted to be 1 s.d. Detections with *fkqual* 4 are considered to be noise.

An F-statistic is calculated to estimate the coherence of the f-k power spectrum:

Algorithms ▼

$$fstat = (J - 1) \cdot \left[\frac{fk_{\max}}{1 - fk_{\max}} \right] \quad (143)$$

where fk_{\max} is the maximum f-k spectrum power normalized to unity by the total power in the J array elements. Compare equation (143) with the F-Statistic presented in equation (116). The value $fstat$ estimated by the f-k spectrum is identical to that estimated in the traditional F-Statistic calculation with the condition that both entities be calculated

- over the same time-duration,
- starting at the same initial time,
- within the same spectral band,
- for the same subset of sensor-channels,
- sampling the vector slowness at the same coordinate points, and
- where the band pass filtered waveform data from each sensor channel have been transformed to zero-mean unit-variance processes.

This statistic can be zero or any positive value.

Slowness and azimuth uncertainties ($delslo$ and $delaz$) that are stored in the **detection** and **arrival** database tables are calculated by using an estimate of the processing band center frequency, f_c , the F-Statistic value, $fstat$, and fk-spectrum processing recipe parameters (dk , for the measurement error, which is site-specific, and ds for the modeling error shown in equation (144)).

$$delslo = \sqrt{(ds)^2 + \left[\frac{(dk)^2}{fstat \cdot f_c^2} \right]} \quad (144)$$

where $delslo$, as calculated in this equation, is in seconds per degree of distance on the surface of the earth. Prior to writing $delslo$ to the **detection** database table $delslo$ is converted to slowness units of s/km (the same as the units for magnitude slowness, $slow$); however, in the **arrival** database table the $delslo$ value is in units of s/deg. as is the $slow$ value in that table.

The F-Statistic is closely related to the coherent $(snr)_e$ through equation (145):

$$fstat = 1 + J \cdot (\text{snr})_e \quad (145)$$

where coherent snr, $(\text{snr})_e$, is the ratio of the spatially coherent signal power to the incoherent noise power in the frequency band. The term $1/f_c^2$ in equation (144) is closely related to the center wavelength, λ_c , of the band via equation (146):

$$\lambda_c = \frac{1}{|\vec{s}| \cdot f_c} \quad (146)$$

where \vec{s} is the magnitude slowness of the plane-wave contribution with frequency f_c . Consequently, the measurement-error contribution to the slowness variance, $(\text{delslo})^2$, can be expressed as equation (147):

$$(\text{delslo})_{\text{meas}}^2 = |\vec{s}|^2 \cdot [\lambda_c]^2 \cdot \frac{1}{J} \cdot \frac{1}{(\text{snr})_e} \cdot (dk)^2 \quad (147)$$

This formulation is similar in structure to the parametric form of the Cramer-Rao estimated slowness variance in equation (150). The factor $(dk)^2$ embodies all of the site-specific geometry of the sensor-array, which is borne by the term $\left[\frac{1}{2\pi}\right]^2 \cdot [\Phi^T \cdot \mathbf{K}_r \cdot \Phi]^{-1}$ in equation (150). The dimensionless factor $\frac{1}{2T(\text{BW})_p}$ is the processing gain of the coherent processing, which for fixed coherent integration time, T , and processing bandwidth, $(\text{BW})_p$, is simply an adjustment to the value of $(\text{snr})_e$.

After the *delso* is estimated with contributions both from the modeling and the measurement errors then azimuth error, *delaz*, can be estimated as follows:

$$\text{delaz} = \begin{cases} 2 \cdot \text{asin}\left(\frac{1}{2} \cdot \frac{\text{delslo}}{\text{slow}}\right); & \frac{\text{delslo}}{\text{slow}} < 2 \\ 180.0; & \frac{\text{delslo}}{\text{slow}} \geq 2 \end{cases} \quad (148)$$

Algorithms ▼

**MODIFIED CRAMER-RAO
VARIANCE ESTIMATION**

Modified Cramer-Rao variance estimation is used for estimating the measurement error in azimuth and slowness for infrasonic signals. It provides the most optimistic estimate of the variance. The virtue of the Cramer-Rao bound is that it parametrically attributes the impact of the relevant processing, signal, and ambient noise parameters to the estimated variance in a rational fashion. In most situations genuine data whose attributes are known (or with synthetically generated data) are sufficient to determine the value of a multiplication factor, K (> 1). K is used to multiply the Cramer-Rao bound to match the attributes of the given estimated parameter and its estimation algorithm.

The estimation of variances for a plane-wave arrival at a spatially diverse sensor system can be accomplished via a calculation that estimates the Cramer-Rao bound (CRB). The assumptions are as follows:

- An ideal plane wave sweeps across a spatially diverse planar array of point sensors.
- The energy in the plane wave is perfectly coherent in a spatial sense.
- The noise present at each sensor site is spatially uncorrelated with respect to the noise at any other sensor site and temporally uncorrelated with the plane wave arrival.

Also, the expectation value of the local noise level at each sensor site is time stationary and equal from site to site. Given this ideal scenario, the CRBs for the variance, σ_ϕ^2 , in the arrival azimuth, ϕ , and the variance, σ_s^2 , in the magnitude of the slowness, $|\hat{s}|$, are given by equations (149) and (150):

$$\sigma_\phi^2 = \left[\frac{180}{\pi} \right]^2 \cdot \left[\frac{\langle \lambda \rangle}{2\pi} \right]^2 \cdot \frac{1}{2T(BW)_p} \cdot \frac{1}{J} \cdot \frac{1}{(snr)_e} \cdot \left[\frac{d\Phi^T}{d\phi} \cdot \underset{\sim}{K}_r \cdot \frac{d}{d\phi} \Phi \right]^{-1} \quad (149)$$

$$\sigma_s^2 = |\hat{s}|^2 \cdot \left[\frac{\langle \lambda \rangle}{2\pi} \right]^2 \cdot \frac{1}{2T(BW)_p} \cdot \frac{1}{J} \cdot \frac{1}{(snr)_e} \cdot [\Phi^T \cdot \underset{\sim}{K}_r \cdot \Phi]^{-1} \quad (150)$$

where 2 X 2 matrix $\underset{\sim}{K}_r$ is given by

$$\underline{K}_r = \left(\frac{1}{J} \cdot \sum_{i=1}^J \underline{r}_i \underline{r}_i^T \right) - \left(\frac{1}{J} \cdot \sum_{i=1}^J \underline{r}_i \right) \cdot w \cdot \left(\frac{1}{J} \cdot \sum_{i=1}^J \underline{r}_i^T \right) \quad (151)$$

The 2 X 1 column matrix, Φ , its transpose, Φ^T , and their derivatives with respect to the azimuth, ϕ , are given by

$$\Phi = \begin{bmatrix} \sin(\phi) \\ \cos(\phi) \end{bmatrix}, \Phi^T = [\sin(\phi) \cos(\phi)], \frac{d\Phi}{d\phi} = \begin{bmatrix} \cos(\phi) \\ -\sin(\phi) \end{bmatrix} \text{ and } \frac{d\Phi^T}{d\phi} = [\cos(\phi) -\sin(\phi)] \quad (152)$$

The 2 X 1 column matrix, \underline{r}_i , its transpose, \underline{r}_i^T , and their outer product, $\underline{r}_i \underline{r}_i^T$, are given by

$$\underline{r}_i = \begin{bmatrix} deast_i \\ dnorth_i \end{bmatrix}, \underline{r}_i^T = [deast_i \ dnorth_i] \text{ and } \underline{r}_i \underline{r}_i^T = \begin{bmatrix} deast_i^2 & deast_i dnorth_i \\ dnorth_i deast_i & dnorth_i^2 \end{bmatrix} \quad (153)$$

where $deast_i$ and $dnorth_i$ (in the **site** database table) are the distances east and north of each sensor element from a reference point for the array station.

The entities $\left[\frac{d\Phi^T}{d\phi} \cdot \underline{K}_r \cdot \frac{d\Phi}{d\phi} \right]$ and $[\Phi^T \cdot \underline{K}_r \cdot \Phi]$ are real-valued scalars that, if not zero valued, can be inverted trivially. The parameter w is a real-valued scalar to be defined below; $(snr)_e$ is the signal-to-noise power ratio (assumed to be the same at all sensor-channels); $(BW)_p$ is the bandwidth of the bandpass filter applied to the raw waveform data; T is the integration time used in generating either the normalized cross-correlation functions or the F-statistic estimates; $|\vec{s}|^2$ is the magnitude-squared of the vector slowness (for the plane-wave arrival); and $\langle \lambda \rangle$ is the “effective” wavelength (in kilometers) for the plane-wave arrival.

The scalar parameter w and the “effective” wavelength originate in the conversion of a “monochromatic” (single wavelength) model of the plane wave into a more complicated model wherein the wave is wideband. A key assumption in this conversion is that the slowness is constant for each frequency component of the wideband process.

Computationally, w is defined by equation (154)

Algorithms ▼

$$w = \frac{\left[\sum_{k=k_1}^{k_2} \frac{|a_k|^2}{\lambda_k^2} \right]}{\sum_{k=k_1}^{k_2} \frac{|a_k|^2}{\lambda_k^2}} \quad (154)$$

and $\langle \lambda \rangle$ is defined by equation (155):

$$\frac{1}{\langle \lambda \rangle^2} = \sum_{k=k_1}^k \frac{|a_k|^2}{\lambda_k^2} \quad (155)$$

where the a_k are complex-valued normalized spectral weights, which can be estimated from the Fourier spectral density function of the bandpass filtered waveform data used to generate either the normalized cross-correlation functions or the estimates of the F-statistic. The index k runs over the frequency-domain spectral pickets from the lowest frequency in the bandpass, defined by picket index k_1 , to the highest frequency in the bandpass, defined by picket index k_2 . The indexed-wavelengths, λ_k , are merely the wavelengths corresponding to the frequencies of the spectral pickets, also indexed by k . The complex-valued normalized spectral weights, a_k , are defined by equation (156):

$$a_k = \frac{p(k)}{\sum_{k=k_1}^{k_2} |p(k)|^2} \quad (156)$$

where the complex-valued unnormalized spectral weights, $p(k)$, are given by the Fourier transformation of the bandpass filtered beamformed time series of time duration defined by M samples (which corresponds to the time duration of the energy packet of the plane-wave arrival) starting with the time-domain sample index, n_0 , (which corresponds to the plane wave's onset time):

$$p(k) = \frac{1}{M} \cdot \sum_{m=0}^{M-1} \left\{ b(m+n_0) \cdot e^{-2\pi\sqrt{-1}(m \cdot k)/M} \right\} \quad (157)$$

where $b(m)$, is the beam formed in the direction and apparent phase slowness of the plane wave arrival.

The squared inverse of the “effective” wavelength is simply the weighted sum of the squared inverses of the wavelengths of the individual Fourier components of the wideband process (within the filter bandpass). The real-valued scalar weight, w , is the ratio of the square of the first moment of inverse wavelength to the second moment of inverse wavelength.

POLARIZATION ANALYSIS

Polarization analysis uses three orthogonal components of motion (usually from 3-C seismometers) to obtain estimates of the size of signals, their rectilinearity, their planarity, and measures of the directions of motion. The algorithms are based on Jurkevics [Jur88] and are implemented in the *DFX* application (see [Wah96a] and [Wah96b]).

To analyze the three orthogonal components of data for polarization, the data are bandpass filtered, and an identical series of overlapping cosine-tapered time windows are applied to the filtered data. The polarization is calculated separately for each window, and the values stored in the database are determined by the size of various polarization features (such as the rectilinearity). The polarization within a time window is calculated as follows:

Let $X = [x_{ij}]$, where i is the index of the sample (ranging from 1 to N) in the window, and j is the index of the component (vertical = 1, north = 2, east = 3). The covariance matrix, S , is evaluated as follows:

$$S_{ik} = \frac{XX^T}{N} = \left[\frac{1}{N} \sum_{i=1}^N x_{ij} x_{ik} \right] \quad (158)$$

The terms of the three-by-three covariance matrix are the auto- and cross-variances of the three components of motion:

$$S = \begin{bmatrix} S_{zz} & S_{nz} & S_{ez} \\ S_{zn} & S_{nn} & S_{en} \\ S_{ze} & S_{ne} & S_{ee} \end{bmatrix} \quad (159)$$

Algorithms ▼

Solving the eigenproblem for S yields the eigenvalues ($\lambda_1, \lambda_2, \lambda_3$) and eigenvectors (u_1, u_2, u_3) for the matrix. These values are used to compute the polarization parameters described in the following paragraphs:

The 3-C amplitude of the signal is computed by summing the square roots of the individual component amplitude as shown in equation (160). Three values are stored in the **apma** database table, *ampp*, *amps*, and *amplr*. *Ampp* is measured in the same window as the maximum rectilinearity; *amps* and *amplr* are the maximum values of 3-C amplitude.

$$ampp, amps, amplr = \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2} \quad (160)$$

Signal rectilinearity is a measure (between 0 and 1) of the linearity of the particle motion as shown in equation (161). This measurement is stored in the **arrival** database table as *rect*.

$$rect = 1 - \frac{\lambda_3 + \lambda_2}{2\lambda_1} \quad (161)$$

Planarity is a measure of the planar characteristic of the polarization ellipsoid, as shown in equation (162). A plane wave has a zero-valued third eigenvalue and a planarity of 1.0. Planarity is stored in the **apma** database table as *plans* (S-phase planarity) and *planlr* (Rayleigh-phase planarity).

$$plans, planlr = 1 - \frac{2\lambda_3}{\lambda_2 + \lambda_1} \quad (162)$$

The long-axis incidence angle is the apparent incidence angle (measured from the vertical) of the eigenvector (u_1) associated with the largest eigenvalue (λ_1), as shown in equation (163). Long-axis incidence angle is stored in the **apma** database table as *inang1*.

$$inang1 = \frac{\arccos(|u_{11}|)}{90} \quad (163)$$

where u_{11} is the direction cosine of the eigenvector associated with the largest eigenvalue.

The short-axis incidence angle is the apparent incidence angle (measured from the vertical) of the eigenvector (\mathbf{u}_3) associated with the smallest eigenvalue (λ_3) as shown in equation (164). Short-axis incidence angle is stored in the **apma** database table as *inang3*.

$$inang3 = \frac{\arccos(|u_{31}|)}{90} \quad (164)$$

where u_{31} is the direction cosine of the eigenvector associated with the smallest eigenvalue.

The horizontal-to-vertical power ratio is computed for two windows: in one window the rectilinearity is maximum (*hvratp*), and in the other window 3-C amplitude is maximum (*hvrat*). See equation (165). Both *hvratp* and *hvrat* are stored in the **apma** database table.

$$hvrat, hvratp = \frac{S_{ee} + S_{nn}}{2S_{zz}} \quad (165)$$

Horizontal-to-vertical power ratios are also computed for several different frequency bands that are one octave wide and centered at frequencies of 0.25, 0.50, 1.00, 2.00, and 4.00 Hz. These ratios are stored in the **amp3c** database table as *htov* with the center frequencies stored in *cfreq*.

The observed azimuth of the signal (assuming that the signal is a P-phase) is the azimuth of the eigenvector (\mathbf{u}_1) associated with the smallest eigenvalue (λ_1):

$$seazp = \text{atan2}\left(\frac{u_{13}}{u_{12}}\right) \quad (166)$$

The observed azimuth of the signal (assuming that the signal is a S-phase) is the azimuth of the eigenvector (\mathbf{u}_3) associated with the smallest eigenvalue (λ_1):

$$seazs = \text{atan2}\left(\frac{u_{33}}{u_{32}}\right) \quad (167)$$

Two polarization parameters (*hmxmn* and *seazlr*) are obtained by solving the 2-D eigensystem using only the horizontal components.

Algorithms ▼

The maximum-to-minimum ratio of the horizontal components is the ratio of the largest eigenvalue (λ_1) to the smallest eigenvalue (λ_2) as shown in equation (168). The result is stored in the **apma** database table as *hmxmn*.

$$hmxmn = \sqrt{\frac{\lambda_1}{\lambda_2}} \quad (168)$$

The Rayleigh-phase observed azimuth of the signal (*seazlr*) is computed from the horizontal eigenvectors:

$$seazlr = \text{atan2}\left(\frac{u_{11}}{u_{21}}\right) \quad (169)$$

CEPSTRAL ANALYSIS

Cepstral analysis is used to measure periodicity of scalloping in the amplitude spectrum of a signal. Scalloping is generally due to multiple coherent signals arriving within the time window used to calculate the spectrum. Scalloping may occur when multiple charges are fired in sequence for a mining explosion or when the bubble created by an in-water explosion successively expands and collapses.

The real cepstrum, $C(\tau)$, as a function of the delay time between successive source pulses, τ , is defined “schematically” as follows:

$$C(\tau) = \text{real}(\text{ifft}\{\log[\|\text{fft}(\text{signal})\|]\}) \quad (170)$$

where *ifft* and *fft* are the inverse and forward Fast Fourier Transform. In the idealized case of coherent signals with constant time delay between arrivals and in the absence of noise, the maximum of the real cepstrum will occur at the delay time between the arriving signals. The cepstrum is sensitive to the same problems that exist in normal Fourier analysis so that many of the steps in computing the cepstrum are similar to power spectrum estimation.

In practice, the algorithm is as follows:

1. Noise, $w(n)$, and signal data, $s(n)$, of equal lengths, N , are retrieved using time windows derived during the initial hydroacoustic feature extraction. A linear regression is computed and subtracted for both waveform segments to remove any DC offset or drift in the time series data. Given a raw time series of waveform data, $d[n]$, (d serving as a symbol for either w or s depending on whether the waveform data are ambient noise or signal, respectively) where n is the time-domain sample index that lies in the range $0 \leq n \leq N-1$ standard least sum of squared errors linear regression, yields the following intercept, d_0 , and slope, σ , regression parameters:

$$d_0 = \frac{\{(4N+2) \cdot T_d\} - \{6 \cdot U_d\}}{\{N \cdot (N-1)\}} \quad (171)$$

$$\sigma = \frac{\{(12 \cdot U_d) - 6 \cdot (N+1) \cdot T_d\}}{\{(N-1) \cdot N \cdot (N+1)\}} \quad (172)$$

where T_d , the sum of the raw $d[n]$ over N , is given by

$$T_d = \sum_{n=0}^{(N-1)} d[n] \quad (173)$$

and U_d , the sum of the raw $n \cdot d[n]$ over N , is given by

$$U_d = \sum_{n=0}^{(N-1)} (n+1) \cdot d[n] \quad (174)$$

and the detrended waveform data time-series, $\bar{d}[n]$, is given by

$$\bar{d}[n] = d[n] - (d_0 + n \cdot \sigma) \quad (175)$$

Algorithms ▼

2. The detrended data, $\bar{d}[n]$, are multiplied by a cosine taper, yielding, $\bar{d}_{\text{tap}}[n]$, to reduce spectral leakage, and a Fourier transformation of $\bar{d}_{\text{tap}}[n]$ is computed to produce the amplitude spectra for both the noise and signal time series, as in

$$\bar{d}_{\text{tap}}[n] = \bar{d}[n] \cdot \frac{1}{2} \cdot \left[1 - \cos\left(\frac{\pi n}{n_0}\right) \right] \quad (176)$$

where n lies in the interval $0 \leq n \leq n_0 - 1$ and

$$\bar{d}_{\text{tap}}[n] = \bar{d}[n] \cdot \frac{1}{2} \cdot \left[1 - \cos\left(\frac{\pi[(N-1)-n]}{n_0}\right) \right] \quad (177)$$

where n lies in the interval $N - n_0 \leq n \leq (N - 1)$, otherwise

$$\bar{d}_{\text{tap}}[n] = \bar{d}[n] \quad (178)$$

where n lies in the interval $n_0 \leq n \leq (N - 1) - n_0$, and n_0 is given by

$$n_0 = [\text{taper} \cdot N] \quad (179)$$

where *taper* is a real positive number of magnitude less than unity, specified in the hydro processing recipe. The Fourier transformation, $D[k]$, of the zero-padded cosine tapered detrended real-valued waveform time-series data are given by

$$D[k] = \begin{bmatrix} T \\ K \end{bmatrix} \cdot \sum_{n=0}^{K-1} \bar{d}'_{\text{tap}}(n) \cdot e^{-2\pi\sqrt{-1}(n \cdot k)/K} \quad (180)$$

where $D[k]$ is the complex-valued amplitude of the k -th frequency picket, and K is the smallest number of form 2^p equal to or greater than N (where p is a positive integer). T is the time span represented by K time-domain samples of the process $\bar{d}'_{\text{tap}}[n]$, which is the zero-padded version of $\bar{d}_{\text{tap}}[n]$, given by

$$\bar{d}'_{\text{tap}}[n] = \bar{d}_{\text{tap}}[n] \quad (181)$$

where $0 \leq n \leq (N - 1)$, and

$$\bar{d}'_{\text{tap}}[n] = 0 \quad (182)$$

where $N \leq n \leq (K-1)$.

3. Frequency smoothing, using a boxcar filter, is applied to reduce the variance in the spectra. The spectra are then scaled by the total energy in the signal so that the magnitude of the computed features will not depend on the original signal strength. The multi-pass smoothing is accomplished by replacing the spectral density function, $|D[k]|^2$, with a smoothed spectral density function, $|\tilde{D}[k]|^2$, which is itself smoothed in an identical fashion, and so forth for typically three or more passes. The number of passes is defined in the hydro processing recipe. The boxcar smoothing algorithm for each single pass is given by

$$|\tilde{D}[k]|^2 = \left[\frac{1}{(k_2 - k_1)} \right] \cdot \sum_{\kappa = k_1}^{k_2 - 1} |D[\kappa]|^2 \quad (183)$$

where the frequency picket index limits, k_1 and k_2 , are set in accordance with

$$k_1 = \text{MAX} \left\{ 0, \left[k - \left(\frac{\Delta k}{2} \right) \right] \right\} \quad (184)$$

and

$$k_2 = \text{MIN} \left\{ K, \left[k + \left(\frac{\Delta k}{2} \right) + 1 \right] \right\} \quad (185)$$

and Δk is the frequency picket width over which the smoothing is performed. Equations (184) and (185) ensure that the beginning and end regions of the Δk are calculated in such fashion as to be an odd-valued positive integer based upon the cepstral “box width” parameter set in the hydro processing recipe. The power in the signal is scaled by the following equation:

Algorithms ▼

$$P_{sig} = \left[\frac{K}{T} \right] \cdot \sum_{k=0}^{(K/2)} |\tilde{D}_{sig}[k]|^2 \quad (186)$$

where P_{sig} is the total power in the positive frequency domain of the multi-pass smoothed signal spectral density function. Then, the multi-pass smoothed power spectral density functions are normalized for both the signal and the noise, $|\tilde{D}_{sig}[k]|^2$ and $|\tilde{D}_{noi}[k]|^2$, respectively:

$$\langle |\tilde{D}_{sig}[k]|^2 \rangle = |\tilde{D}_{sig}|^2 / P_{sig} \quad (187)$$

$$\langle |\tilde{D}_{noi}[k]|^2 \rangle = |\tilde{D}_{noi}|^2 / P_{sig} \quad (188)$$

to obtain the normalized, multi-pass smoothed power spectral density functions: $\langle |\tilde{D}_{sig}[k]|^2 \rangle$ and $\langle |\tilde{D}_{noi}[k]|^2 \rangle$, for both the signal and the noise.

4. The noise spectrum is subtracted from the signal spectrum to remove any scalloping present in the background noise:

$$\Psi[k] = \langle |\tilde{D}_{sig}[k]|^2 \rangle - \langle |\tilde{D}_{noi}[k]|^2 \rangle \quad (189)$$

This spectrum, $\psi[k]$, is separated into two parts: a low-pass filtered version (or trend—obtained by internal parameter *nflag* set to 1) and the detrended spectrum (obtained by internal parameter *nflag* set to 0). In both cases, the spectrum is first band-pass filtered in accordance with the frequency limits, *flo* and *fhi*, which are defined in the hydro processing recipe. The low-pass filtered version of the spectrum is useful for finding short time delay signals, which produce broad scalloping in the frequency domain, and the detrended spectrum is most sensitive to long time delay signals. Results from the cepstrum of the “trend” spectrum are saved in the hydro features with the string “trend” in the parameter name. Results from the cepstrum of the detrended spectrum are saved in the hydro features with the string “signal” in the parameter name. In either case, before the inverse Fourier transform is applied either to the

“trend” version of $\psi[k]$ or to the detrended version of $\psi[k]$, several adjustments are applied to $\psi[k]$. For the “trend” spectrum approach, $\psi[k]$ must be adjusted as follows:

1. Shift $\psi[k]$ to be positive valued over the pickets klo through khi corresponding to flo and fhi .
2. Apply a multi-pass split symmetric window NSE (Noise Spectrum Equalization) to the resulting $\psi[k]$ and replace the in-band values of $\psi[k]$ with the estimated broadband noise background underlying those pickets wherever a narrow bandwidth feature (a “line” component) was present in $\psi[k]$.
3. Set the resulting $\psi[k]$ to the constant value at $\psi[klo]$ for pickets from 0 to klo and for those from khi to $K/2$.
4. Remove the mean of the resulting $\psi[k]$ over the pickets 0– $K/2$ to yield a zero-mean process for inverse Fourier transformation.

For the detrended spectrum approach, $\psi[k]$ must be adjusted as follows:

1. Shift $\psi[k]$ to be positive valued over the pickets klo through khi corresponding to flo and fhi .
2. Apply a multi-pass split symmetric window NSE (Noise Spectrum Equalization) to the resulting $\psi[k]$ thereby estimating the broadband noise background, $B[k]$.
3. Normalize $\psi[k]$ with this estimated underlying broadband noise background, $B[k]$. This process tends to “whiten” the spectrum except in narrow bandwidth regions containing “line” components. With properly chosen NSE parameters (averaging bandwidth, guard bandwidth, thresholds for peak shear, and number of passes) rather than narrow bandwidth line components, the NSE can emphasize spectral lobes of a modulated broadband process.
4. Remove the mean of the resulting $\psi[k]$ over the pickets 0– $K/2$ to yield a zero-mean process for inverse Fourier transformation.
5. Calculate the logarithm of this adjustment

$$\psi[k] = \log_{10}[\text{MAX}(|\Psi[k]|, 0.00000001)] \quad (190)$$

Algorithms ▼

6. Calculate the inverse FFT of both the “trend” and detrended spectrum:

$$C[n] = \text{Re} \left\{ [K] \cdot \sum_{k=0}^{K-1} \psi[k] \cdot e^{2\pi\sqrt{-1}(n \cdot k)/K} \right\} \quad (191)$$

The real part of the inverse Fourier transformation of $\psi[k]$ is the cepstrum, $C(\tau)$.

7. Remove any low-frequency trend in the cepstrum via the same NSE (see [“Noise Spectrum Equalization” on page 225](#)) algorithm used to adjust the spectrum; however, the parameters of the NSE are dramatically different from those applied to the spectrum. In both cases the NSE parameters are set in the hydro processing recipe. By NSEing the cepstrum the peaks stand out. Use the τ corresponding to the largest value in the cepstrum, within some bounding delay time window, as the delay time for the signal. Calculate the variance of the cepstrum over the delay time window. Calculate the mean and standard deviation as follows:

$$\mu = \frac{1}{N} \cdot \sum_{i=K}^M C_i \quad (192)$$

$$\sigma = \sqrt{\frac{1}{N} \cdot \sum_{i=K}^M C_i^2 - \mu^2} \quad (193)$$

where K and M correspond to the minimum and maximum delay times searched, and $N = M - K + 1$. The level of the cepstral peak is measured as the ratio of the magnitude over the standard deviation.

NOISE SPECTRUM EQUALIZATION

Noise Spectrum Equalization, NSE, is a process of estimating the underlying broadband noise energy in a spectrum, which may have relatively narrow high amplitude features. A typical NSE involves:

- application of an averaging window, which is “notched” with a notch width somewhat larger than the width of the narrow high amplitude features to be ignored in estimating the underlying broadband energy
- application of multiple passes over the original spectrum wherein the high amplitude narrow bandwidth features detected in a given pass are replaced by the average underlying broadband energy estimate available in that pass
- reduction of the detection threshold (for high amplitude narrow bandwidth features) from pass to pass by 3 dB levels

Consider an original power spectral density function $P[k]$ defined over picket indices in the closed interval $[0, \bar{K}]$. First copy this array of numbers to a temporary array $Q[k]$. The heart of the algorithm is the application of a notched averaging window of total picket width equal to $\Delta k + (2\Delta g + 1) + \Delta k$. This width is always an odd-valued number because the notch width $(2\Delta g + 1)$ is odd, and the symmetric averaging windows must sum to an even picket count. The underlying broadband energy estimate is simply, $S[k]$, given by the sum of two averages: a lower window, $L[k]$, (“below” the notch) and an upper window, $U[k]$, (“above” the notch), as in

$$S[k] = \frac{\{L[k] + U[k]\}}{\{(\kappa_1 - \kappa_0 + 1) + (\kappa_3 - \kappa_2 + 1)\}} \quad (194)$$

where the lower and upper windows are defined by

$$L[k] = \sum_{k=\kappa_0}^{\kappa_1} Q[k] \quad (195)$$

where $k > \Delta g$, (otherwise 0), and

Algorithms ▼

$$U[k] = \sum_{k=\kappa_2}^{\kappa_3} Q[k] \quad (196)$$

where $k < [\bar{K} - \Delta g]$ (otherwise 0). The picket limits, κ_i , of the two averages (but one facet of the algorithm's complexity) are given by the following equations:

$$\kappa_0 = \text{MAX}[0, k - (\Delta g + \Delta k)] \quad (197)$$

$$\kappa_1 = k - \Delta g \quad (198)$$

$$\kappa_2 = k + (\Delta g + 1) \quad (199)$$

$$\kappa_3 = \text{MIN}\{\bar{K}, [k + (\Delta k + \Delta g + 1)]\} \quad (200)$$

After the quantity $S[k]$ has been generated for k in the closed interval $[0, \bar{K}]$, it can be applied in the “peak shearing” removal of above threshold narrow bandwidth features from the temporary spectrum, as in

$$Q[k] = S[k] \quad (201)$$

whenever $Q[k] > \text{threshold_factor} * S[k]$; otherwise, $Q[k] = P[k]$ (the original spectrum). Where the NSE calls for more than a single pass, the calculations reflected in Equations (195) through (201) are simply repeated with the $Q[k]$'s, which were generated in the previous pass input to the calculations for the current pass. Each time a new pass completes with the “peak shearing” re-evaluation of $Q[k]$, the *threshold_factor* is reduced by a factor of 2 (3 dB).

After the requisite number of passes, typically three, $S[k]$, the broadband underlying energy estimate output from the last pass through the NSE algorithm, can now be used in either of two ways (governed by a flag *noise*). When flag *noise* is set to unity, the original spectrum, $P[k]$, can be replaced by this broadband underlying noise estimate, $S[k]$, as in $P[k] = S[k]$, for all pickets in the closed interval $[0, \bar{K}]$. Or, when flag *noise* is non-unity, the original spectrum, $P[k]$, can be normalized by the broadband underlying noise estimate, $S[k]$, as in: $P[k] = P[k] / S[k]$, for all pickets in the closed interval $[0, \bar{K}]$.

This description was predicated on the input array being a power spectrum, $P[k]$; yet any real-valued, positive array of data can be input. If the *noise* flag is set to unity, then on output the algorithm will replace $P[k]$ with an estimate of the underlying smooth running trend averaged over $2\Delta k$ locations—free of the influence of large “spikes” or transients, which are narrower than $(2\Delta g+1)$ locations. If the *noise* flag is set to non-unity, then on output the algorithm will replace $P[k]$ with a whitened (to near-unity level) version of $P[k]$ in locations containing no large “spikes” or transients, which are narrower than $(2\Delta g+1)$ locations. In locations containing such narrow features, the output value will be the ratio of the narrow feature magnitude to the estimated underlying level at that location. Consequently, the NSE can be applied to a cepstrum or to a correlation function (either auto or cross) to seek out the peak features and normalize them such that the background, in areas where there are no peaks, is near unity.

PROBABILITY WEIGHTED TIME

The probability weighted time is an estimation of the peak energy arrival time. The method can be applied to any signal type, but currently is only applied to hydroacoustic data to estimate arrival times of the peak energy.

Peak Probability

Consider a set of samples from a time series (x_1, \dots, x_N) , which have values $(\lambda_1, \dots, \lambda_N)$ and a sampling interval Δt . The time corresponding to the maximum value of the series is to be determined. However, each value has an uncertainty associated with it, and thus the maximum of λ_i may not be the maximum of x_i . Assume that each sample has the same uncertainty, σ . Then, if the errors are Gaussian distributed, the probability that the true value of x_i lies in some range of width dx_i is

$$P(x_i)dx_i = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i - \lambda_i)^2 / (2\sigma^2)} dx_i \quad (202)$$

The probability that x_i is less than some value, y , given its measured value, λ_i , is

Algorithms ▼

$$P(x_i < y) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^y e^{(x_i - \lambda_i)^2 / (2\sigma^2)} dx_i = \text{erf}\left(\frac{y - \lambda_i}{\sigma}\right) \quad (203)$$

From this, the probability that a particular x_n is the true peak is calculated as follows:

$$P_n \equiv P(x_n \text{ is the peak}) = \int_{\substack{\text{(all possible} \\ \text{values of } x_n)}} \left(\text{probability of any} \right. \\ \left. \text{particular value } x_n \right) \times \prod_{\substack{\text{(all other } x_i \\ i \neq n)}} \left(\text{probability} \right. \\ \left. \text{that } x_i < x_n \right) \quad (204)$$

$$P_n = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dx_n e^{(x_n - \lambda_n)^2 / (2\sigma^2)} \prod_{i \neq n} \text{erf}\left(\frac{x_n - \lambda_i}{\sigma}\right) \quad (205)$$

The behavior of equation (205) becomes more intuitive by substituting variables: $x = (x_n - \lambda_n)/\sigma$.

$$P_n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx e^{-x^2/2} \prod_{i \neq n} \text{erf}\left(x + \frac{\lambda_n - \lambda_i}{\sigma}\right) \quad (206)$$

The second term in the argument of the erf is the number of standard deviations by which two measured values differ. If $\lambda_j < \lambda_k$, the erf in P_j will have lower arguments than the corresponding erf in P_k , and hence $P_j < P_k$. Also, if some λ_k exists such that $(\lambda_k \gg \lambda_j)$ for $j \neq k$, the arguments in the erf for P_k are positive, and the values of the erf approach 1. For all other P_j , the arguments in the erf are negative, and their values approach 0. Hence, $P_k \rightarrow 0$, and all other $P_j \rightarrow 0$.

Arrival-time Estimation

To obtain the probability function for the time series, an estimate of the uncertainty is needed for equation (206). Currently, in hydroacoustic processing the average pre-event noise level is used as the uncertainty in measured values. To estimate the most likely time of arrival of peak energy, a time average of $P(t_i)$ is computed

$$\begin{aligned}
 \text{Probability weighted time} &= \langle t \rangle_P = \sum_n P_n t_n \\
 &= t_0 + \Delta t \sum_n P_n n
 \end{aligned}
 \tag{207}$$

If two samples are near to one another in intensity, the corresponding P_i will be nearly equal, and so $\langle t \rangle_P$ will be an average between the samples. Likewise, if sample j is clearly dominant, $P_j > P_i$ and so $\langle t \rangle_P \sim t_j$.

The variance in the time estimate is

$$\sigma_{\langle t \rangle_P}^2 = \sum_n (P_n t_n^2 - \langle t \rangle_P^2) \tag{208}$$

In the case with a high snr (small uncertainty), if any one sample i dominates, then $P_i \rightarrow 1$, and the measurement error is very small. By contrast, for low snr, the values, P_n , are more evenly distributed, and the measurement error grows large. Identical waveforms may have different measurement errors if they have different noise levels.

NEURAL NETWORKS

Often, detections must be classified into phase types (P, S, and so on) based on a set of features measured from the signal. The number of potential discriminates can be numerous making direct inference methods impractical. However, artificial neural networks provide an alternative approach. Their use in classification type problems is well established, and much literature on the subject is available [\[Bis95\]](#). The basic task of the neural network is to determine the relative likelihoods that a detection belongs to each of the possible categories. For a neural network to do this, it must first be trained. A data set with previously measured features and known phase types is used to train the neural network. After training, the neural network can take new data and compute a number, which represents some measure of the level of membership a detection has in each possible phase category. A decision method is then employed to select between the phase types for this detection or declare the phase type undetermined.

Algorithms ▼

The neural networks used in station processing are known as multi-layer feed-forward networks. The network consists of a set of “nodes” arranged in layers as seen in Figure 34. Features measured from the signal are entered at the input nodes, which pass the features to each of the nodes in the next layer. The output nodes represent the different possible phase types for a detection, that is, categories. The nodes between the input and output nodes have no direct physical interpretation and do not interact outside of the network. These nodes are called hidden layers. In a feed-forward network the output of a node in one layer is transmitted to nodes in the next layer and so forth until the output nodes are reached.

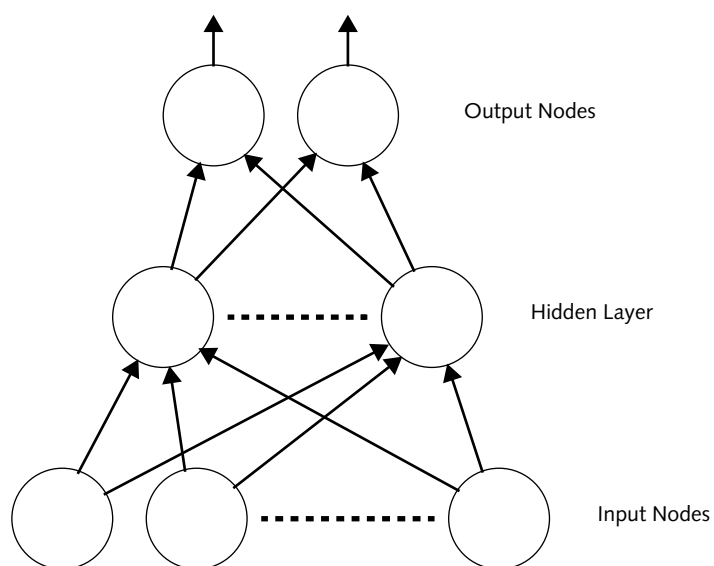


FIGURE 34. NEURAL NETWORK

Each node (with the exception of the input nodes) computes a weighted sum over its input values. This is schematically shown in Figure 35. A parameter called the bias is subtracted from the sum, and then the value is passed through a non-linear “activation” function. The two most common activation functions for continuous

valued input parameters are the sigmoid function, $1/(1 + e^{-a})$, and the hyperbolic tangent. Training the neural network consists of determining the appropriate weights and bias for each node.

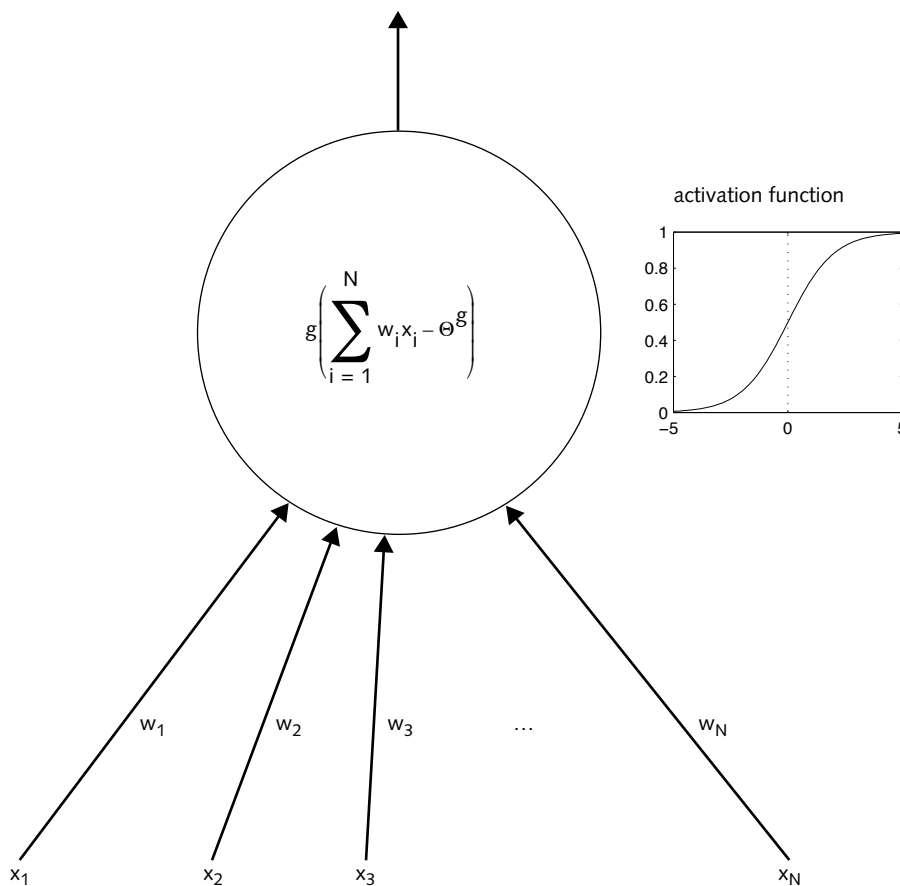


FIGURE 35. INDIVIDUAL NODE IN NEURAL NETWORK

Training the neural network is accomplished with the so called back-propagation method. This algorithm minimizes an error function, typically the rms misfit of the training set, by adjusting the weights of the individual nodes based on the partial derivatives of the error function with respect to the weights. The partial deriva-

Algorithms ▼

tives can be computed for feed-forward networks that have continuously differentiable activation functions. The partial derivatives for the weights at a node can be computed from the derivatives in the layers above. The algorithm computes the partial derivatives in the output layer and then for the next layer down and so forth, hence the name back-propagation. The derivatives are computed and the weights are adjusted in an iterative process until they converge on a solution. Constraints are imposed to avoid over-training (a condition where the network is tuned too specifically to the training set), which can degrade performance with new data. This situation is analogous to overfitting data with an underconstrained model and similar methods of regularization are employed to avoid it.

The study of artificial neural networks is a vast field. Many details are to be considered when creating a neural network such as the number of hidden layers, the number of nodes, the types of activation functions, and the error functions used in training to name a few. See ["References" on page 257](#) for literature containing a more in-depth description of both the theory and implementation of neural networks.

BAYESIAN INFERENCE

A Bayesian method is used to empirically determine the final phase identification. This method is employed in the station processing module called *StaPro*, where final phases of regional S arrivals are identified for single stations. This section describes the method and how to generate the Bayesian inference flat file.

The Bayesian method describes the probability of "a" occurring given conditions of "b" and "c." The desired conditional probability used by *StaPro* is formulated as follows:

$$P(\text{phase_id}|\text{features, context}) = \frac{P(\text{features}|\text{phase_id, context}) \cdot P(\text{phase_id}|\text{context})}{P(\text{features}|\text{context})} \quad (209)$$

Equation (209) describes the probability of "*phase_id*" given the features and context specific to the arrival in question. The right-hand side of the equation can be empirically derived from a ground-truth data set.

Regional S phases (for example, Sn, Lg, Rg, Sx) are the only "*phase_id*'s" used in this method. The "*features*" in the above equation are measurements, made either directly or indirectly by DFX, that describe an arrivals attribute. These features include: S minus P time multiplied by frequency (dimensionless), velocity (km/s), and horizontal-to-vertical power ratio (dimensionless). The "*context*" of an arrival is represented as station specific and includes a description of the arrival pattern (for example, only-S, first-S-of-two, first-S-of-many, or largest-S).

A ground-truth database containing arrival and feature information is used to empirically derive the conditional probabilities for each regional S phase recorded at a station. The probabilities are generated for each phase as a function of the range of features and context situations.

Conditional probabilities for four stations (ARCES, NORES, FINES, and GERES) have been generated and exist in the current flat file called **bays.tbl**. A mechanism in the *StaPro* software allows a default station to be used for a station that does not have station specific probabilities. Currently, this default station is set as NORES.

Newer stations may be added to the flat file with their own probabilities, or a copy of an existing probability from another station located in a similar tectonic setting may be used.

MAGNITUDES

Traditional methods for estimating the magnitude of an event that is detected at a number of spatially diverse stations in a sensor network make use of the magnitude estimates from the detecting stations only. This can lead to over estimation of the event magnitude. In addition to the estimates from detecting stations, Ringdal [Rin76] uses the non-detecting stations by means of a simple statistical model, which calls for the magnitudes at these stations to be below a certain threshold

Algorithms ▼

value. Maximum likelihood estimation is then applied to estimate event magnitudes from the estimates at the detecting stations and the absence of detection at other stations.

The bootstrap procedure of Efron [Efr79] and [Efr81] is applied by McLaughlin [McL88] together with a maximum likelihood estimation (MLE) procedure for estimation event magnitudes to estimate the magnitude of the event and its variance. The bootstrap method establishes a random selection process in which a randomly ordered subset of events is selected and a mean and variance are developed from individual MLE estimates of each of the selected events.

Estimating Body Wave Magnitudes

Body wave magnitudes (m_b) are calculated for terrestrial events from the time-defining primary body waves from seismic stations that are between 20 and 105 degrees from the event.

Station Body Wave Magnitudes

Station magnitudes, $m_b(\text{sta})$, are calculated with the equation:

$$m_b(\text{sta}) = \log\left(\frac{\text{amp}}{\text{per}}\right) + B(\Delta, h) \quad (210)$$

where sta is the station, amp is the half peak-to-peak amplitude ($\text{amplitude.amptype} = A5/2$) of the phase in nm from the **amplitude** database table, per is the period of the largest snr signal in seconds from the **amplitude** database table, and B is an attenuation correction function of distance in degrees (Δ) and event depth (h).

The attenuation corrections as a function of distance and depth are based on [Vei72]. The corrections are tabulated every degree for distances out to 180 degrees and for depths 0, 15, 40 km, and 100–800 km in steps of 100 km. Bi-cubic splines were used for interpolating the tables. The tabulated values were

adjusted for the fact that the original [Vei72](#) tables relate to peak-to-peak amplitudes, whereas the measured amplitudes for m_b calculations are half peak-to-peak.

Station m_b are stored in the **stamag** database table as *magnitude* and identified by the attribute *magtype* = *mb_ave*.

Network Average Body Wave Magnitudes

Network average body wave magnitudes, $m_b(\text{net})$, are formed as averages over station magnitudes, $m_b(\text{sta}_i)$, at seismic stations:

$$m_b(\text{net}) = \frac{1}{nsta} \sum_{i=1}^{nsta} m_b(\text{sta}_i) \quad (211)$$

where *nsta* is the number of stations with magnitudes, $m_b(\text{sta}_i)$, available (and stored as **netmag.nsta**). Outlying station magnitudes deviating by more than one unit of magnitude from the average are removed, and the average is recalculated.

Network average m_b are stored in the **netmag** database table as *magnitude* (defined by the attribute *magtype* = *mb_ave*). The *uncertainty* of a network average m_b in the database table **netmag** is calculated as the standard deviation of the network average magnitude, $m_b(\text{net})$, if the number of stations, *nsta*, is two or more:

$$uncertainty = \frac{1}{nsta - 1} \sqrt{\sum_1 [m_b(\text{sta}_i) - m_b(\text{net})]^2} \quad (212)$$

Maximum Likelihood Body Wave Magnitudes

In addition to the amplitude and period measurements from recognized seismic phases, maximum likelihood estimates (MLE) of m_b incorporate amplitude and period measurements from stations that do not have recognized arrivals for the

Algorithms ▼

event. The magnitude estimates are made using arrival data from primary seismic stations and assumed standard deviations of station magnitudes of 0.35 magnitude units.

The probability of detection, $P(\text{Detect}|\mu, \sigma)$, of an event with true magnitude μ is given by the following equation:

$$P(\text{Detect}|\mu, \sigma) = P(m \geq a|\mu, \sigma) = \Phi\left[\frac{(\mu - a)}{\sigma}\right] \quad (213)$$

where $\Phi()$ is the standard cumulative normal distribution function, and μ and σ are the mean and standard deviation of a Gaussian distribution, m is the observed station magnitude, and a is the threshold setting. (For any given event, the observed magnitude at the individual stations of a worldwide homogeneous seismic network is modeled to follow a Gaussian distribution.) In this case, the likelihood of observing the set of n station magnitudes $\{m_1, m_2, \dots, m_i, \dots, m_n\}$ is given by the following equation:

$$L(m_1 \dots m_n|\mu, \sigma) = \left\{ \prod_{i \in \{D\}} \frac{1}{\sigma} \cdot \phi\left[\frac{(m_i - \mu)}{\sigma}\right] \right\} \cdot \left\{ \prod_{j \notin \{D\}} \frac{1}{\sigma} \cdot \phi\left[\frac{(a_j - \mu)}{\sigma}\right] \right\} \quad (214)$$

where $\{D\}$ is the set of detecting stations; a_j is the threshold at station j , which was in the set of nondetecting stations; and $m_i > a_i$ for $i \in D$. There are q non-detecting stations and $m_j \leq a_j$ for $j \notin D$. ϕ is the standard Gaussian probability density function. Given a set of observed m 's and known a 's, and the assumed value of σ ($= 0.35$), a maximization of L with respect to μ will yield an estimate of the true value of the event magnitude in light of the subsets of $(n-q)$ detecting and q non-detecting stations.

Maximum likelihood m_b are stored in the **netmag** database table as *magnitude* (defined by *magtype* = *mb_mle*). The uncertainty of MLE network magnitudes are obtained with bootstrapping [McL88].

In the bootstrap approach a set of N observations is first used to estimate the population mean and standard deviation: (μ_1, σ_1) . Then N data are selected in random order from the set of N data without regard to whether a given datum had

been selected previously. Thus, N (with replacement) has N^N combinations of random selections of N objects, which is a very large number even when N is modest: 10^{10} . These resampled data are used to estimate (μ_2, σ_2) . This random selection and resampling procedure is applied M times (where M is a moderate number: 20); and the bootstrap estimate of the population mean, m , is the arithmetic mean of $\{m_1, m_2, \dots, m_M\}$; and the population standard deviation, $\bar{\sigma}$, is the arithmetic mean of $\{\sigma_1, \sigma_2, \dots, \sigma_M\}$. The uncertainty in m is the standard error of the mean of the resampled estimates $\{m_1, m_2, \dots, m_M\}$. This latter uncertainty is stored in the **netmag** table.

Estimating Local Magnitudes

Local magnitudes are calculated by using time-defining P and P_n phases recorded at seismic stations less than 20 degrees from the event. Magnitudes are calculated only for shallow events according to the following rules:

- If the depth error is not estimated, M_L is calculated only if the depth is less than or equal to 40 km.
- If a depth error is available, M_L is calculated only if the probability of the depth being more than 40 km is less than 0.9.

M_L calculations are limited to P and P_n phases. The intent with using only P phases is to make M_L consistent with m_b and make it less dependent on source type.

Station Local Magnitudes

A station local magnitude, $M_L(\text{sta})$, is calculated according to the equation:

$$M_L(\text{sta}) = 0.5 \cdot \log(stav^2 - ltav^2) + A(\text{distance}) \quad (215)$$

where sta is the station, $stav$ is the short-term average of the amplitude from the **detection** database table, $ltav$ is the long-term average of the amplitude and is computed as $ltav = \text{detection.stav} / \text{detection.snr}$, distance is the distance from the event to the station in kilometers, and A is an attenuation correction that is a function of distance.

Algorithms ▼

The attenuation correction, A , can be tailored to each station for each phase. The corrections are stored in parameter files and were derived from data accumulated during GSETT-3. No correction for focal depth is applied.

Station local magnitudes are stored in the IDC database table **stamag** as *magnitude* (defined by the attribute, *magtype* = m1ppn).

Network Average Local Magnitudes

A network local magnitude is formed from the station magnitudes, which are summed for all contributing stations in a weighted scheme with a local magnitude uncertainty, σ_i , assigned to each station depending on the phase:

$$M_L(\text{net}) = \frac{\sum_i \frac{M_L(\text{sta}_i)}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}} \quad (216)$$

where stations that have default values for the attenuation correction, A , in the station magnitude are assigned a σ_i of 0.60, and other stations are assigned a σ_i of 0.45. These standard deviations were established based on the scatter of the data used for normalizing M_L to m_b .

Outlying station local magnitudes deviating by more than one unit of magnitude from the weighted average are removed, and the weighted average is recalculated.

Network local magnitudes are stored in the IDC database table **netmag** as *magnitude* (defined by the attribute, *magtype* = m1ppn). The *uncertainty* of the network local magnitude in the IDC database table, **netmag**, is calculated from the standard deviations of the station magnitudes, σ_i :

$$\text{uncertainty} = \frac{1}{\sqrt{\sum_i \frac{1}{\sigma_i^2}}} \quad (217)$$

Estimating Surface Wave Magnitudes

Station Surface Wave Magnitudes

Station surface wave magnitude, $M_s(sta)$ is calculated according to the equation:

$$M_s(sta) = \log(amp/per) + B(\Delta) \quad (218)$$

where *amp* and *per* are from the **amplitude** table (*amptype* = ALR/2, or ANL/2). ALR/2 identifies Rayleigh wave signal measurements, and ANL/2 identifies noise measurements. *A* is the maximum amplitude/period within the allowed period range: 18 to 22 seconds. $B(\Delta)$ (equation (219)) is an attenuation formula recommended by [Rez88].

$$B(\Delta) = \left[\frac{1}{3} \cdot \log(\Delta) \right] + \left\{ \frac{1}{2} \cdot \log[\sin(\Delta)] \right\} + [0.0046 \cdot \Delta] + 2.370 \quad (219)$$

where the term 0.0046 is a γ times $\log(e)$ ($= 0.4343$), and γ , the attenuation coefficient, was determined empirically to be 0.0105 using a very large data set.

Station M_s are stored in the **stamag** database table as *magnitude* and identified by the attribute *magtype* = ms_ave.

Network Average Surface Wave Magnitudes

Network average surface wave magnitudes, $M_s(net)$, are formed as averages over $M_s(sta_i)$, at primary seismic stations:

$$M_s(net) = \frac{1}{nsta} \sum_{i=1}^{nsta} M_s(sta_i) \quad (220)$$

where *nsta* is the number of stations with station magnitudes (and stored in the **netmag** database table).

Algorithms ▼

Outlying station magnitudes deviating by more than one unit of magnitude from the average are removed, and the average is recalculated.

The network average M_s are stored in the IDC database table **netmag** as *magnitude* (defined by *magtype* = *ms_ave*). The *uncertainty* of a network average M_s magnitude in the IDC database table, **netmag**, is calculated as the standard deviation of the network magnitude if the number of stations, *nsta*, is two or more:

$$uncertainty = \frac{1}{nsta - 1} \sqrt{\sum_i [M_s(sta_i) - M_s(net)]^2} \quad (221)$$

Maximum Likelihood Surface Wave Magnitudes

Maximum likelihood M_s are estimated (*magtype* = *ms_mle*) as described for m_b . The estimates are based on data from primary seismic stations and assumed standard deviations of station magnitudes of 0.25 magnitude units. The uncertainty of MLE network magnitudes are, as for the MLE of m_b , obtained with bootstrapping. In cases where no signal amplitudes have been recorded at stations between 2 and 100 degrees, upper bound magnitudes are calculated as the magnitude at which the probability of no detecting stations between 2 and 100 degrees is 0.95, assuming a Gaussian distribution.

DETECTION PROBABILITY

The probability of detection is a function of the estimated amplitude (which is determined differently depending on whether the station is at regional or teleseismic distances), the station noise level, signal and noise amplitude variance, and a user-supplied snr threshold for detection.

The probability that an event will be detected at a station is given by the equation:

$$\text{Prob}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{\left(-\frac{y^2}{2}\right)} dy \quad (222)$$

where

$$y = \frac{\log(\text{Amp}) - \log(\text{Noise}) - \log(\text{snr-thresh})}{\sqrt{\sigma_{\text{amp}}^2 + \sigma_{\text{noise}}^2}} \quad (223)$$

The $\log(\text{Amp})$ term is a function of the distance between the event and the station. If the station is at a regional distance from the event location, then amplitude is estimated using the following relation:

$$\log(\text{Amp}) = M - A(R) \quad (224)$$

where M is the original M_L ; A is an attenuation correction, and R is the station-event distance in km. If an M_L measurement is unavailable then, m_b or M_s is used. If no valid magnitude is available, a user-supplied default is used. Noise is obtained from the **siteaux** table. The snr-thresh is the user-supplied snr detection threshold, which can be different for each station. σ_{amp}^2 is the estimated variance of log amplitude, and σ_{noise}^2 is the variance of log noise from the **siteaux** table. Because a station may have multiple **siteaux** entries, the noise and σ_{noise}^2 values are selected as the entries with the shortest period within a user-supplied period range.

The variance in the amplitude, σ_{amp}^2 , is an empirically developed estimate from a very large database of events measured world wide. The current value for σ_{amp} is 0.35.

In the case where station-event distances are teleseismic, the value for Amp is found via the relation:

$$\log(\text{Amp}) = M - B(\text{delta}, \text{depth}) \quad (225)$$

where M is the m_b (if m_b is unavailable M_L or M_s is used), and $B(\text{delta}, \text{depth})$ is a distance and depth correction value obtained from an attenuation file [\[Vei72\]](#). In the teleseismic case, σ_{amp} is also obtained from the attenuation file (see documentation for *libprob* for details on the teleseismic probability of detection calculation methods and the attenuation data file). The noise and σ_{noise}^2 are obtained from the **siteaux** table and are selected as the entries with the shortest period within the user-supplied period range.

Algorithms ▼

The detection probability is used in the context of a likelihood function where the likelihood, L , of an event, given observed magnitudes $\{m_1, \dots, m_i, \dots, m_n\}$ at n stations, wherein $(n-q)$ detect and q do not detect, is

$$\log\{[L(m_1 \dots m_i \dots m_n)]\} = \left\{ \sum_{i \in \{n-q\}} \{\log(P_i)\} \right\} + \left\{ \sum_{j \in \{q\}} \{\log(1 - P_j)\} \right\} \quad (226)$$

where $\{n-q\}$ is the set of $n-q$ detecting stations, and $\{q\}$ is the set of q nondetecting stations, and where P_i is the probability of detection of the signal at the i^{th} station. The expected value of $\log(L)$, $\langle \log(L) \rangle$, is given by

$$\log(L) = \sum_{i \in \{n\}} \{[P_i \cdot \log(P_i)] + [(1 - P_i) \cdot \log(1 - P_i)]\} \quad (227)$$

where $\{n\}$ is the entire set of n stations. The quantity, $\log(L)$, is closely related to the entropy of the system. The variance of this measure is given by

$$\text{VAR} = \sum_{i \in \{n\}} [P_i(1 - P_i)] \cdot [\log(P_i) - \log(1 - P_i)]^2 \quad (228)$$

The standard deviation, σ , can be taken as the square root of VAR. The $\log(L)$ value can now be compared with its expected value, $\langle \log(L) \rangle$, in view of the variance, and if the deviation is too large the association of this group is deemed to be unacceptable, as in

$$x = \frac{[\log(L) - \langle \log(L) \rangle]}{\sigma} \quad (229)$$

That is, if $x > x_0$, where the threshold for comparison is x_0 , then the association is unacceptable.

The detection probability, P , at a given station is given by

$$P = \frac{1}{\sqrt{2\pi}} \cdot \int_{-y}^{\infty} e^{-u^2/2} du \quad (230)$$

where lower limit, $-y$, is obtained from the quantity

$$y = \frac{\{snr - \log(snr_0)\}}{\sqrt{[(sstdv)^2 + (nstdv)^2]}} \quad (231)$$

where snr is the distance and depth-corrected magnitude at that station; snr_0 is the threshold snr required for detection at that station; $sstdv$ is the signal standard deviation in log amplitude units; and $nstdv$ is the noise standard deviation at that station also in log amplitude units. The value of snr is derived by

$$snr = event_mag - [B(\Delta) - 0.028\sqrt{z}] - \log[noise_i] \quad (232)$$

where $event_mag$ is the event magnitude (m_b if available, otherwise M_L); B is the attenuation correction as a function of distance; Δ , in degrees; z is the event depth in kilometers; and $noise_i$ is the station noise at the i th station in nanometers.

EVENT SCREENING STATISTICS

Event screening is based on the statistical treatment of the event characteristics and their uncertainties. The statistics used for event screening based on depth and relative body and surface wave magnitudes are presented in this section.

Depth

Given the depth estimate, \hat{D} , and its variance, σ_D^2 , the distribution of the square of the depth estimate (normalized to zero mean and unit variance) is obtained as the marginal distribution of the hypocenter estimate with respect to origin time and epicenter. This distribution is chi-squared with one degree of freedom, which is equivalent to a normal distribution for the depth estimate. That is, $(\hat{D} - D)/\sigma_D$ has a normal distribution with zero mean and unit variance, where D is the true (unknown) depth. Under this assumption, a one-sided $100(1 - \alpha)\%$ confidence interval for the depth, D , is $[\hat{D} - z_\alpha \sigma_D, \infty)$, where z_α is the α -percentile of the normal distribution with zero mean and unit variance. That is, z_α is defined such that $P[z > z_\alpha] = 1 - \Phi(z_\alpha) = \alpha$, where

Algorithms ▼

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z dx e^{-x^2/2} = 1 - \frac{1}{\sqrt{2\pi}} \int_z^{\infty} dx e^{-x^2/2} \quad (233)$$

An event is screened out at the $100(1 - \alpha)\%$ confidence level if the shallowest part of the depth confidence interval is deeper than the depth threshold D_T :

$$D - z_\alpha \sigma_D > D_T \quad (234)$$

The *default* depth threshold is $D_T = 10$ km. For a *default* confidence level of 97.5%, $z_\alpha = 1.96$.

Alternate values of the depth threshold, D_T , and the confidence level may be specified in a custom event-screening request or subscription.

The depth screening criterion is equivalent to performing a test of hypothesis at the α significance level, that the true depth, D , is less than or equal to a threshold depth D_T . The hypothesis is rejected if the inequality in (234) is satisfied.

A “score” for the depth screening criterion is defined as

$$\text{score}_{\text{depth}} = \frac{\hat{D} - D_T}{z_\alpha \sigma_D} - 1 \quad (235)$$

A positive *score* corresponds to a one-sided $100(1 - \alpha)\%$ confidence interval for the true depth that is greater than D_T ; and, hence, the inequality in (234) is satisfied. For a default 97.5% confidence interval, the *score* increments approximately by one unit for each two-sigma interval of depth estimate from the depth threshold.

In this analysis the errors are assumed to be distributed according to a Gaussian (normal) distribution for positive depths, with zero probability for values above the ground (negative depth estimates are constrained to the surface.) Also, the data variance is assumed to be independent of depth and the value of σ_D^2 used for the variance is assumed to sufficiently represent the true uncertainty.

Surface Wave Magnitude:Body Wave Magnitude

For an event for which both network m_b and M_s are estimated, the difference between m_b and M_s is calculated as $Am_b - M_s$, where A is a constant slope term that accounts for possible magnitude dependence of the $M_s:m_b$ relation.

The standard deviations used for the individual m_b and M_s measurements are $\sigma_b = \sigma_s = 0.3$, based on an uncertainty analysis performed by [Bot96]. The variance of the network average of N_b independent measurements of m_b is given by σ_b^2/N_b , with a similar expression for M_s . Assuming the m_b and M_s estimates are uncorrelated, the variance of $Am_b - M_s$ is

$$\sigma_M^2 = \text{var}(Am_b - M_s) = A^2 \frac{\sigma_b^2}{N_b} + \frac{\sigma_s^2}{N_s} \quad (236)$$

Assuming that $Am_b - M_s$ has a normal distribution with variance σ_M^2 , a one-sided $100(1 - \alpha)\%$ confidence interval for $Am_b - M_s$ is $(-\infty, Am_b - M_s + z_\alpha \sigma_M]$, where z_α is the α -percentile of the normal distribution with zero mean and unit variance, as defined in equation (233). An event is “screened out” if this confidence interval is entirely less than a threshold, M_T (in other words, if it is entirely outside of the explosion population). Thus, an event is “screened out” if

$$Am_b - M_s + z_\alpha \sigma_M < M_T \quad (237)$$

This screening criterion is equivalent to performing a test of hypothesis at the α significance level that the true value of $Am_b - M_s$ is greater than or equal to a threshold, M_T . The hypothesis is rejected if the inequality in equation (237) is satisfied.

[Mar72] have suggested using a threshold of $m_b - M_s = 1.2$, based on historical explosion and earthquake data. Based on careful calibration of PIDC magnitudes relative to historical data, [Mur97], the Technical Experts Group has recommended that $A = 1.25$ and $M_T = 2.20$ be used [WGB98]. Other values of the slope, A , the threshold, M_T , and the confidence level may be specified in a custom event-screening request or subscription.

Algorithms ▼

A “score” for the $M_s:m_b$ screening criterion is defined as

$$\text{score}_{\text{mbms}} = \frac{M_T - Am_b + M_s}{z_\alpha \sigma_M} - 1 \quad (238)$$

A positive score corresponds to a one-sided $100(1 - \alpha)\%$ confidence interval for $Am_b - M_s$ entirely less than M_T . For the default 97.5% confidence level, the score increments approximately one unit for each two-sigma interval from the screening threshold.

Combined Screening Score

A combined screening score is computed to indicate numerically the degree to which an event does, or does not, meet the composite depth and $M_s:m_b$ screening criteria. To define the combined score, let

$$x = \frac{\hat{D} - D_T}{\sigma_D} \quad (239)$$

and

$$y = \frac{M_T - Am_b + M_s}{\sigma_M} \quad (240)$$

Under the null hypothesis that $D = D_T$ and $Am_b - M_s = M_T$, x and y both have normal distributions with zero mean and unit variance, and x and y are statistically independent. Thus, $z = (x + y)/\sqrt{2}$ also has a normal distribution with zero mean and unit variance.

As for the individual scores defined for the depth and $M_s:m_b$ screening criteria, the combined score must be normalized such that a score of zero corresponds to a $100(1 - \alpha)\%$ confidence level that an explosion at a maximum depth of D_T would not be screened out. This corresponds to $z < z_\alpha$, where z_α is the α -percentile of the normal distribution with zero mean and unit variance, as defined in equation [\(233\)](#).

Thus, the combined score for a $100(1 - \alpha)\%$ confidence level, based on the depth and $M_s:m_b$ screening criteria, is given by

$$\text{score}_{\text{combined}} = z/z_\alpha - 1 = \frac{1}{\sqrt{2}} \left[\frac{\hat{D} - D_T}{z_\alpha \sigma_D} + \frac{M_T - A m_b + M_s}{z_\alpha \sigma_M} \right] - 1 \quad (241)$$

Alternative values of the slope, A , the thresholds, D_T and M_T , and the confidence level may be specified in a custom event-screening request or subscription.

For cases where only depth or $M_s:m_b$ is available for an event, the combined score is given by the appropriate individual score defined in equations (235) and (238).

AUTOMATIC ASSOCIATION CHI-SQUARED TEST

This test is used within the grid-based association process in *GA* to screen corroborating arrivals after a driver arrival has been identified. A two-station location is performed using the available attributes of the driver and candidate corroborating arrivals, and the sum of squares of the normalized residuals are compared to the chi-square threshold for the number of degrees of freedom in the problem and the desired level of confidence. If the sum is larger than the threshold, the corroborating arrival is not added to the association set seeded by the driver arrival. The confidence level is adjustable but usually set at 0.99, meaning that a valid corroborating arrival has a 1% chance of being rejected.

Mathematical Formalism

The basic hypothesis being tested is as follows: given a grid location, a *driver* arrival valid for this grid cell location, and a potential corroborating arrival:

“Can the current corroborating arrival be associated with the driver arrival for an event within the cell (with some level of confidence)?”

The least-squares problem that is solved using a QR decomposition of the A matrix is formalized by the following equation:

Algorithms ▼

$$(A^T A) \Delta \hat{M} = A^T \Delta d \quad (1) \quad (242)$$

The $\Delta \hat{M}$ vector contains the origin time t , location of event with respect to the grid point (Δx , Δy , Δz), and the magnitude of the event. Matrix A is the Frechet kernel (partial derivatives) matrix at the center of the cell. The Δd vector is the difference between the observed data and the estimated data at the center of the cell. The data vector also contains the location of the event with respect to the center of the cell. The Δd vector is the difference between the observed data and the estimated data at the center of the cell. The data vector also contains the location of the event with respect to the center of the cell. This ensures a stable solution to the least-squares problem and is equivalent to a damping of the linear inverse problem. The relationship between model and data vectors is explicitly written in the coordinate system defined with the x axis joining the grid point to the driver station for the case when all data elements for the two arrivals are available:

$$\Delta \hat{M} = A \Delta d \quad (243)$$

where

$$\Delta d = \begin{bmatrix} \Delta T \\ \Delta m \\ \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \quad (244)$$

and

$$A = \begin{bmatrix} 1 & 0 & \left(\frac{\partial}{\partial r}t(r1)\right) & 0 & \left(\frac{\partial}{\partial z}t(r1)\right) \\ 1 & 0 & \left(\frac{\partial}{\partial r}t(r2)\cos\theta\right) & \left(\frac{\partial}{\partial r}t(r2)\sin\theta\right) & \left(\frac{\partial}{\partial z}t(r2)\right) \\ 0 & 1 & \left(\frac{\partial}{\partial r}a(r1)\right) & 0 & \left(\frac{\partial}{\partial z}a(r1)\right) \\ 0 & 1 & \left(\frac{\partial}{\partial r}a(r2)\cos\theta\right) & \left(\frac{\partial}{\partial r}a(r2)\sin\theta\right) & \left(\frac{\partial}{\partial z}a(r2)\right) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \left(\frac{|p1|}{\sin r1}\right) & 0 \\ 0 & 0 & \left(\frac{|p2|\sin^2\theta}{\sin r2}\right) & \left(\frac{-|p2|\sin\theta\cos\theta}{\sin r2}\right) & 0 \\ 0 & 0 & \left(\frac{-|p2|\sin\theta\cos\theta}{\sin r2}\right) & \left(\frac{|p2|\cos^2\theta}{\sin r2}\right) & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (245)$$

then:

$$\Delta \hat{\mathbf{M}} = \begin{bmatrix} \Delta t1 \\ \Delta t2 \\ \Delta a1 \\ \Delta a2 \\ \Delta p1_x \\ \Delta p1_y \\ \Delta p2_x \\ \Delta p2_y \\ \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \quad (246)$$

Algorithms ▼

Index 1 refers to the elements for the driver station and arrival whereas index 2 refers to the corroborating station elements. The Δ 's in the data vector refer to differences between observations and expected values for a location at the center of the cell. Δt_1 and Δt_2 are the travel-time residuals, Δa_1 and Δa_2 the amplitude residuals, \mathbf{p}_1 and \mathbf{p}_2 are the slowness vectors, $|\mathbf{p}_1|$ and $|\mathbf{p}_2|$ are their moduli, and $\Delta \mathbf{p}_1$ and $\Delta \mathbf{p}_2$ are the slowness vector residuals. The partial derivatives in the A matrix are for the time and amplitude functions $a(r)$ and $t(r)$. The geometry of the problem is laid out in [Figure 36](#). The distances from the center of the cell to the driver station and corroborating station are respectively r_1 and r_2 . Note that the Δd vector has the peculiarity that some of the components corresponding to the driver arrival will be zero because the estimated origin time, magnitude, and location are derived from the driver arrival. The slowness vector components of the matrix are computed to first order, assuming no lateral variation of the slowness and no first-order variation of the slowness with depth. In addition, a small angle approximation is made in the derivation of the corresponding A matrix elements. This approximation implies that the test should not be applied when the small angle approximation is not valid. Whether or not the test is applied is controlled by one of the parameters in the automatic association process.

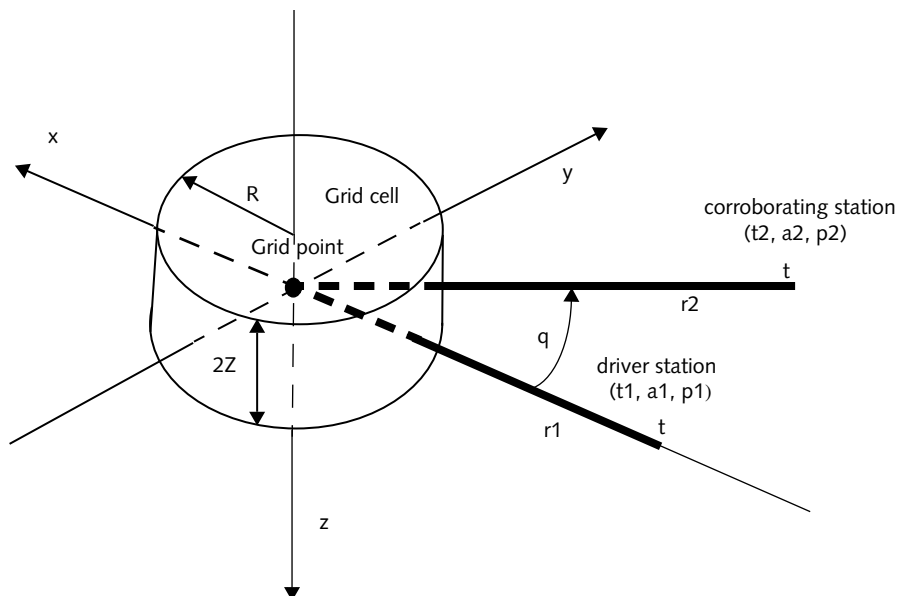


FIGURE 36. GEOMETRY FOR AUTOMATIC ASSOCIATION CHI-SQUARE TEST

Chi-square Test

Σ_d is the diagonal matrix with elements $1/\sigma_{t1}^2$, $1/\sigma_{t2}^2$, $1/\sigma_a^2$, $1/\sigma_a^2$, $1/\sigma_{p1}^2$, $1/\sigma_{p1}^2$, $1/\sigma_{p2}^2$, $1/\sigma_{p2}^2$, $1/R^2$, $1/R^2$, $1/Z^2$; the test compares the quantity $\Delta d^T \Sigma_d \Delta d$ to the chi-square distribution value $\chi(p, \nu)$, where ν is the number of degrees of freedom (number of data vector elements minus number of model vector element), and p is the confidence (or probability) level. The confidence level is usually set to 0.99.

Algorithms ▼

The σ_t 's are equal to the time standard deviation (measurement + modeling); σ_a is equal to the amplitude measurement standard deviation; the σ_p 's are the measurement standard deviation in slowness; R is the effective radius of the beam cell; and Z is the effective half thickness of the cell.

EVENT LOCATION ERROR ESTIMATION

The model covariance matrix, C_m , from the inversion of data for event location can be viewed as an effective measure of the network geometry (a fixed shape) scaled by the applied weights. Ideally these weights reflect the physical knowledge as a magnitude scaled onto a fixed unweighted geometry. Weighting can be applied, assuming no prior knowledge, thereby relying purely on the data residuals from the previous iteration. Alternatively, the chosen weighting can assume exact prior knowledge based on uncertainty determinations from past studies (these take the form of measurement and modeling errors). Lastly, a combination of both prior and current knowledge may be exploited together. Regardless, none of these approaches should change the general character of the covariance matrix, only its magnitude.

Assuming convergence has been reached in the location process, an error ellipse geometry can be prescribed as points about the final location. At a given p percentile confidence level, each ellipse point, $\mathbf{x}_{\text{ellip}}$, is defined on an ellipsoid about the event location, $\mathbf{x}_{\text{final}}$, as

$$(\mathbf{x}_{\text{ellip}} - \mathbf{x}_{\text{final}})^T C_m (\mathbf{x}_{\text{ellip}} - \mathbf{x}_{\text{final}}) = \kappa_p^2 \quad (247)$$

where the confidence coefficient, κ_p^2 , can be represented as

$$\kappa_p^2 = M \bar{s}^2 F_p[M, N - M] \quad (248)$$

with F_p defining a p percentile F-statistic distribution with M and $N - M$ degrees of freedom (N is the number of data available, and M is the number of parameters to be determined). \bar{s}^2 is the variance scale factor and can be interpreted as the mean

ratio of the actual versus assumed data variances. Flinn [\[Fli65\]](#) originally constructed this scale factor based on a normalized sample variance, s_s^2 , exclusively the domain of the actual data residuals as

$$s_s^2 = \frac{|\mathbf{r}_w|^2}{N - M} \quad (249)$$

where \mathbf{r}_w is the normalized residual vector. This traditional strategy is employed to compute error ellipses in programs such as HYPOINVERSE [\[Kle78\]](#). Such error ellipses are referred to as “confidence ellipses.” These ellipses are strictly a function of the final set of data residuals computed when convergence has been reached. No prior knowledge is incorporated in the error estimation process. Unfortunately, when few data are available, a small numerator or denominator in the above equation can cause an unreasonably large or small confidence ellipse to be computed. For an even-determined problem where all data residuals are 0.0, the confidence ellipse collapses to a point. In practice, the other case is more common, that is, the small denominator results in excessively large confidence ellipses. The quadratic approximation employed to calculate the error ellipsoid relies heavily on the assumption of linearity. The error ellipse is itself an approximation of a more complicated two-dimensional space. The true error spaces are not quite as simple, especially for small events.

Evernden [\[Eve69\]](#) approached this problem from the opposite point of view. He surmised that one could develop good prior knowledge about the data uncertainties. With this knowledge one could determine more representative error ellipses. In essence, one assumes exact prior errors can be determined. Evernden showed that this approach was most practical for events with few defining associations. As such, the confidence coefficient, κ_p^2 , is reduced to a p percentile χ^2 distribution with M degrees of freedom, $\chi_p^2[M]$. An error ellipse computed assuming exact prior knowledge is a “coverage ellipse”.

Jordan and Sverdrup [\[Jor81\]](#) took this logic one step further by recasting the confidence coefficient in terms of both prior and current information, namely,

$$\kappa_p^2 = M s^{-2} F_p[M, K + N - M] \quad (250)$$

Algorithms ▼

which permits the current estimate of the variance scale factor to be represented as

$$s_s^2 = \frac{Ks_K^2 + |\mathbf{r}_w|^2}{K + N - M}, \quad (251)$$

where, s_K^2 , is a prior estimate for the variance scale factor, \bar{s}^2 . When $K = 0$, the equation is reduced to that of Flinn [\[Fli65\]](#). When $K = \text{infinity}$, then the data variances are assumed to be known exactly, the normalized sample variance equals 1, and the equation is reduced to the χ^2 solution of Evernden [\[Eve69\]](#). By exploiting this Bayesian approach, intermediate values of K can be developed to best reflect the available imperfect prior knowledge and at the same time down-weight the importance of the current data residuals. Ideally, the K should be made as large as possible to remove the influence of the current information altogether. Error ellipses that result from intermediate values of K are referred to as “ K -weighted error ellipses.” The resultant error ellipsoid is a function of prior estimates and current assumptions about the random data characteristics.

By refining the prior knowledge as more experience is gained, K approaches infinity, and a χ^2 error (coverage) ellipsoid results. To achieve this goal two primary types of prior errors should be distinguished: measurement error, and modeling error. These errors are combined as a total rms error whose inverse is the weighting applied (multiplied) to the associated row of sensitivity matrix, \mathbf{A} , and residual vector, \mathbf{r} . The measurement error is simply the uncertainty in the measure itself. In the case of an arrival time, the measurement error is the estimated uncertainty in the pick (largely, a function of the snr). The modeling error is the uncertainty attached to the assumed model used to determine a given theoretical measure (for example, an uncertainty assigned to a one-dimensional travel-time table at a given distance/depth based on studies undertaken to develop the travel-time table in the first place).

The error ellipses are calculated from the appropriate marginal variances following an F-distribution. These ellipses are normalized in the sense that they are scaled to a particular confidence probability, thereby making the marginal variances justified. The following parameters are described:

- two-dimensional confidence error ellipse of the epicenter, as determined from the 2 x 2 covariance matrix of the epicenter (marginal with respect to origin time and depth)
- one-dimensional confidence intervals of origin time and focal depth, as determined from the scalar variance of the origin time and focal depth, respectively (marginal with respect to origin time/focal depth and epicenter)

The diagonal elements of the data resolution matrix, which can be constructed as part of the error estimation process, are the data importances [\[Min74\]](#). The data importances indicate how important a given datum is to the overall solution. For the over (even)-determined problem defined here, an individual data importance is a scalar between 0.0 and 1.0, and the sum of all data importances are equal to the rank of the matrix. For a standard epicentral location the rank is 3 (latitude, longitude, and origin time) and by adding depth to the solution the rank increases to 4. These data importances can be extremely useful for an analyst attempting to assess how valuable a given datum is to their solution.

References

The following sources are referenced in or supplement this document:

- [Bac90] Bache, T. C., Bratt, S. R., Wang, J., Fung, R. M., Kobryn, C., and Given, J., "The Intelligent Monitoring System," *Bull. Seism. Soc. Am.*, 80, Part B, 1833-1851, 1990.
- [Bac93] Bache, T. C., Bratt, S. R., Swanger, H. J., Beall, G. W., and Dashiell, F. K., "Knowledge-Based Interpretation of Seismic Data in the Intelligent Monitoring System," *Bull. Seism. Soc. Am.*, 83, 1507-1526, 1993.
- [Bau88] Baumgardt, D. R. and Zeigler, K. A., "Spectral Evidence for Source Multiplicity in Explosions: Applications to Regional Discrimination of Earthquakes and Explosions", *Bull. Seis. Soc. Am.*, 78, 1773-1775, 1988.
- [Bis95] Bishop, C. M., *Neural Networks for Pattern Recognition*, Oxford University Press Inc., New York, 1995.
- [Bon98] Bondar, I., *Teleseismic Slowness-azimuth Station Corrections (SASC) for the IMS Network*, CCB-PRO-98/01, 59 pp., 1998. (online) http://www.pidc.org/librarybox/lib_top.html
- [Bot96] Bottone, S., Fisk, M. D., Gray, H. L., and McCartor, G. D., *The Dependence of Magnitude Uncertainty on Station Coverage*, PL-TR-96-2250, Phillips Laboratory, Hanscom AFB, MA, U.S.A., 1996.
- [Bra88] Bratt, S. R. and Bache, T. C., "Locating Events with a Sparse Network of Regional Arrays," *Bull. Seism. Soc. Am.* 78, 780-798, 1988.

References ▼

- [Coy96] Coyne, J., and Henson, I., *Geotool Sourcebook*, CMR-96/06, 1996.
- [Dzi76] Dziewonski, A., and Gilbert, F., "The Effect of Small, Aspherical Perturbations on Travel Times and a Re-examination of the Correction for Ellipticity," *Geophys. J. R. astr. Soc.*, 44, 7-17, 1976.
- [Efr79] Efron, B., "Bootstrap Methods, Another Look at the Jackknife," *Ann. Statistics*, 7, 1-26, 1979.
- [Efr81] Efron, B., "Censored Data and the Bootstrap," *J. Am. Statist. Assoc.*, 76, 312-319, 1981.
- [Eve69] Evernden, J., "Precision of Epicenters Obtained by Small Numbers of World-wide Stations," *Bull. Seism. Soc. Am.*, 59, 1365-1398, 1969.
- [Fli65] Flinn, E., "Confidence Regions and Error Determinations for Seismic Event Location," *Rev. Geophys.*, 3, 157-185, 1965.
- [Gei10] Geiger, L., "Herdbestimmung bei Erdbeben ans den Ankunftszeiten," *K. Gessel. Wiss. Goett.*, 4, 331-349, 1910.
- [Gol89] Golub, G. and VanLoan, C., *Matrix Computations*, 2nd Ed., The Johns Hopkins University Press, Baltimore/London, 1989.
- [GSE91] Group of Scientific Experts, *Progress Report to the Conference of Disarmament on the 42nd Session of the Ad Hoc Group of Scientific Experts to Consider International Co-operative Measures to Detect and Identify Seismic Events*, CRP/214, 1991.
- [GSE94] Group of Scientific Experts, *Progress Report to the Conference of Disarmament on the 42nd Session of the Ad Hoc Group of Scientific Experts to Consider International Co-operative Measures to Detect and Identify Seismic Events*, CRP/259, 1994.
- [GSE95b] Group of Scientific Experts, *GSETT 3 Documentation, Volume Two: Operations*, CRP/243, 1995.
(online) http://www.pidc.org/librarybox/lib_top.html

- [Hed89] Hedlin, M. A. H., Minster, J.B., and Orcutt, J. A., "The Time-frequency Characteristics of Quarry Blasts and Calibration Explosions Recorded in Kazakhstan, U.S.S.R.," *Geophys. J. Int.*, 99, 109-121, 1989.
- [Hed90] Hedlin, M. A. H., Minster, J.B., and Orcutt, J. A., "An Automatic Means to Discriminate between Earthquakes and Quarry Blasts," *Bull. Seism. Soc. Am.*, 80, 2143-2160, 1990.
- [IDC5.1.1Rev1] Science Applications International Corporation, Pacific-Sierra Research Corporation, *Database Schema (Part 1, Part 2, and Part 3), Revision 1*, SAIC-99/3009, PSR-99/TN1142, 1999.
- [IDC6.2.5] Science Applications International Corporation, *Analyst Instructions for Seismic, Hydroacoustic, and Infrasonic Data*, SAIC-98/3002, 1998.
- [IDC6.4Rev1] Science Applications International Corporation, *IDC Software Man Pages (Library Functions, User Commands), Revision 1*, SAIC-99/3033, 1999.
- [IDC7.1.2] Science Applications International Corporation, *Station Processing* (in preparation).
- [IDC7.2.2] Science Applications International Corporation, *Analyst Review Station Scheme Functions*, SAIC-99/3026, 1999.
- [Jen92] Jenkins, R., Nagy, W., Sereno, T., Swanger, H., and Wahl, D., "Regional Attenuation and Travel-time Characteristics at GERESS," *EOS Trans., Am. Geophys. Un.*, 73, 374, 1992.
- [Jor81] Jordan, T., and Sverdrup, "Teleseismic Location Techniques and their Application to Earthquake Clusters in the South-central Pacific," *Bull. Seism. Soc. Am.*, 71, 1105-1130, 1981.
- [Jur88] Jurkevics, A., "Polarization Analysis of Three-component Data," *Bull. Seism. Soc. Am.* 78, 1725-1743, 1988.
- [Kam92] Kamigaichi, O., *A Fully Automated Method for Determining the Arrival Times of Seismic Waves and Its Application to an On-line Processing System*, GSE/Japan/40, Paper tabled in the 34th GSE session in Geneva, July, 1992.

References ▼

- [Kam94] Kamigaichi, O., *Automated Identification of Arrival Time, etc., Using AR-model*, Paper presented at the GSE workshop in Tokyo, Japan 14–16 March 1994.
- [Ken91a] Kennett, B., *IASPEI 1991 Seismological Tables*, Research School of Earth Sciences, Australian National University, 167 pp., 1991.
- [Ken91b] Kennett, B., and Engdahl, E., "Traveltimes for Global Earthquake Location and Phase Identification," *Geophys. J. Int.*, 105, 429-465, 1991.
- [Ken95] Kennett, B., "Event Location and Source Characterization," *Monitoring a Comprehensive Test Ban Treaty, NATO ASI Series, Series E: Applied Sciences*, Vol. 303, 501-520, eds. E. Husebye and A. Dainty, Kluwer Academic Publishers, Dordrecht, 1995.
- [Kle78] Klein, F., "Hypocenter Location Program HYPOINVERSE," *U.S. Geol. Surv., Open-File Rept. 78-694*, 113 pp., 1978.
- [Kvæ86] Kværna, T., and Doornbos, D. J., "An Integrated Approach to Slowness Analysis with Arrays and Three-component Stations," *Semiannual Technical Summary, 1 October 1985-31 March 1986*, NORSAR Sci. Rept. 1-86/87, Kjeller, Norway, 1986.
- [Kvæ89] Kværna, T., "On Exploitation of Small-aperture NORESS-type Arrays for Enhanced P-wave Detectability," *Bull. Seism. Soc. Am.* 79, 888–900, 1989.
- [Kvæ95] Kværna, T., "Automatic Onset Time Estimation Based on Autoregressive Processing," *Semiannual Technical Summary, 1 April–30 September 1995*, NORSAR Sci. Report 1–95/96, Kjeller, Norway, 113–133, 1995.
- [LeB96a] LeBras, R., *User Manual for the Global Association Subsystem (GA)*, Science Applications International Corporation, SAIC-96/1128, 1996.
- [LeB96b] LeBras, R., *Programmer's Guide for the Global Association Subsystem (GA)*, Science Applications International Corporation, SAIC-96/1129, 1996.

- [Mar72] Marshall, P. D., and Basham, P. W., "Discriminating between Earthquakes and Underground Explosions Employing an Improved M_s Scale," *Geophys. J. Roy. Astron. Soc.*, 28, 431-458, 1972.
- [McL88] McLaughlin, K., "Maximum Likelihood Event Magnitude Estimation with Bootstrapping for Uncertainty Estimation," *Bull. Seism. Soc. Am.*, 78, 855-862, 1988.
- [Min74] Minster, J. B., Jordan, T. H., Molnar, P., and Haines, E., "Numerical Modelling of Instantaneous Plate Tectonics," *Geophys. J.R. Astr. Soc.*, 36, 541-576, 1974.
- [Mur97] Murphy, J. R., "Calibration of IMS Magnitudes for Event Screening Using the M_s/m_b Criterion," *Proceedings of the Event Screening Workshop: Beijing, China*, November 4-7, 1997.
- [Myk84] Mykkeltveit, S., and Bungum, H., "Processing of Regional Seismic Events Using Data from Small Aperture Arrays," *Bull. Seism. Soc. Am.* 74, 2313-2333, 1984
- [Myk90] Mykkeltveit, S., Ringdal, F., Kværna, T., and Alewine, R., "Application of Regional Arrays in Seismic Verification Research," *Bull. Seism. Soc. Am.* 80, 1777-1800, 1990.
- [Nag95] Nagy, W. C., *AFTAC Distributed Subsurface Network EvLoc Design and Maintenance*, Science Applications International Corporation, SAIC-95/1153, 1995.
- [Nag96] Nagy, W., *New Region-dependent Travel-time Handling Facilities at the IDC; Functionality, Testing, and Implementation Details*, Technical Report SAIC-96/1179, 57 pp., 1996.
- [Pai82] Paige, C., and Saunders, M., "LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares," *ACM Trans. Math. Soft.*, 8, 43-71, 1982.
- [Pre88] Press, W., Flannery, B., Teukolsky, S., and Vetterling, W., *Numerical Recipes in C, The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1988.

References ▼

- [Rez88] Rezapour, M., and Pearce, R.G., "Bias in Surface-wave Magnitude M_s Due to Inadequate Distance Correction," *Bull. Seism. Soc. Am.*, 88, 43-61, 1988.
- [Rin76] Ringdal, F., "Maximum Likelihood Estimation of Seismic Magnitude," *Bull. Seism. Soc. Am.*, 66, 789-802, 1976.
- [Rin89] Ringdal, F., and Kværna, T., "A Multi-channel Processing Approach to Real-time Network Detection, Phase Association, and Threshold Monitoring," *Bull. Seism. Soc. Am.*, 79, 780-798, 1989.
- [Rin90] Ringdal, F., Editor, "Regional Seismic Arrays and Nuclear Test Ban Verification," Special Symposia Issue, *Bull. Seism. Soc. Am.*, 80, Part B, 1775-2281, 1990.
- [Sai96] Saikia, C. K., Thio, H. K., Woods, B. B., Song, X., Zhu, L., and Helmberger, D. V., "Path Calibration, Source Estimation, and Regional Discrimination for the Middle East and Western Mediterranean," PL-TR-96-2307, Phillips Laboratory, Hanscom AFB, MA, 1996.
- [Ser92] Sereno, T., Swanger, H., Jenkins, R., Nagy, W., and Wahl, D., "Attenuation and Travel-time Characteristics of Regional Phases Recorded at GERESS," *Symposium on Regional Seismic Arrays*, 22-24 June, Bavaria, Germany, 1992.
- [Ser93] Sereno, T. J., and Patnaik, G. B., *Initial Wave-type Identification with Neural Networks and Its Contribution to Automated Processing in IMS Version 3.0*, Science Applications International Corporation, SAIC-93/1219, 1993.
- [Ste97] Stevens, J. L., and McLaughlin, K. L., *Improved Methods for Regionalized Surface Wave Analysis*, Final Report PL-TR-97-2135, Phillips Laboratory, Directorate of Geophysics, AFMC, Hanscom AFB, MA, 01731-3010, 1997.
- [Sut91] Suteau-Henson, A., "Three-component Analysis of Regional Phases at NORESS and ARCESS: Polarization and Phase Identification," *Bull. Seism. Soc. Am.* 81, 2419-2440, 1991.

▼ References

- [Vei72] Veith, K. F., and Clawson, G. E., "Magnitude of Short-period P Wave Data," *Bull. Seism. Soc. Am.*, 62, 435-453, 1972.
- [Vei75] Veith, K., "Refined Hypocenters and Accurate Reliability Estimates," *Bull. Seism. Soc. Am.*, 65, 1199-1222, 1975.
- [Wah96a] Wahl, D. D., *Programmer's Guide for the Detection and Feature Extraction Program*, Science Applications International Corporation, SAIC-96/1069, 1996.
- [Wah96b] Wahl, D. D., *User's Manual for the Detection and Feature Extraction Program*, Science Applications International Corporation, SAIC-96/1098, 1996.
- [Wan96] Wang, J., *Analyst Review Station User's Manual*, Science Applications International Corporation, SAIC-96/1100, 1996.
- [Wan97] Wang, J., Israelsson, H., and North, R. G., *Adaptive Training Approach to Neural Networks for Seismic Phase Identification*, CMR-97/29, Center for Monitoring Research, 1997.
- [Wei71] Weichert, D. H., "Short-period Spectral Discriminant for Earthquake and Explosion Differentiation," *Z. Geophys.*, 37, 147-152, 1971.
- [WGB98] Working Group B, *Report on the IDC Technical Experts Meeting on Event Screening 5-9 June 1998*, Preparatory Commission of the CTBT, CTBT/WGB-6/TL-2/10, 1998.

Glossary

Symbols

μPa

MicroPascal.

A

AIC

Akaike Information Criterion.

amplitude

Zero-to-peak height of a waveform.

array

Collection of sensors distributed over a finite area (usually in a cross or concentric pattern) and referred to as a single station.

arrival

Signal that has been associated to an event. First, the Global Association (GA) software associates the signal to an event. Later during interactive processing, many arrivals are confirmed and improved by visual inspection.

ARS

Analyst Review Station. This application provides tools for a human analyst to refine and improve the event bulletin by interactive analysis.

associate

Assign an arrival to an event.

associated phase

Phase that is associated with an S/H/I event.

attribute

(1) Characteristic of an item; specifically, a quantitative measure of a S/H/I arrival such as azimuth, slowness, period, and amplitude. (2) A database column.

azimuth

Direction, in degrees clockwise with respect to North, from a station to an event.

B

b value

Slope of the line fit to a plot of seismic magnitude versus cumulative number of events, usually computed for a finite geographic area.

back azimuth

Direction, in degrees, from an event or seismic signal to the station.

Glossary ▼

beam

Waveform created from array station elements that are sequentially summed in the direction of a specified azimuth and slowness.

bulletin

Chronological listing of event origins spanning an interval of time. Often, the specification of each origin or event is accompanied by the event's arrivals and sometimes with the event's waveforms.

C**cepstrum**

Fourier transformation of a power spectrum whose magnitudes have been scaled logarithmically.

channel

Component of motion or distinct stream of data.

coda

Signal of a given phase, which follows the initial waveform of that phase.

coda phase

Detection found within the envelope of a single phase; an otherwise unidentified phase of unknown path designated as tx, Px, or Sx.

coherent

Quality of having a fixed phase relationship; as signals from a wavefront detected on numerous seismic or infrasonic array station elements.

coherent beam

Summation of data from numerous seismic or infrasonic array station elements after shifting the data traces in time to maximize the coherence of plane-wave signals travelling along a particular azimuth and slowness.

component

(1) One dimension of a three-dimensional signal; (2) The vertically or horizontally oriented (north or east) sensor of a station used to measure the dimension; (3) One of the parts of a system; also referred to as a module or unit.

Configuration Control Board

Organizational body that approves and releases new versions of software.

count(s)

Units of digital waveform data.

CRB

Cramer-Rao bound.

crossover

Distance from an event where two different phases arrive at the same time, allowing constructive interference that sometimes enhances the signal amplitudes.

CTBT

Comprehensive Nuclear Test-Ban Treaty (the Treaty).

CTBTO

Comprehensive Nuclear Test-Ban Treaty Organization.

D**DACS**

Distributed Application Control System. This software supports inter-application message passing and process management.

data flow

Sequence in which data are transferred, used, and transformed during the execution of a computer program.

dB

Decibel.

defining

Arrival attribute, such as arrival time, azimuth, or slowness, which is used in calculating the event's location or magnitude.

defining phase

Associated phase for which features are used in the estimation of the location and origin time of an S/H/I event.

deg.

Degrees (as a distance).

detection

Probable signal that has been automatically detected by the *Detection and Feature Extraction (DFX)* software.

DFX

Detection and Feature Extraction.

discard

Action of rejecting real or false seismic events that are insufficiently defined according to the PIDC's rules and guidelines.

discriminant

Characteristic of a seismic event that indicates if the event originated from an earthquake, mining blast, nuclear detonation, or elsewhere.

E**element**

Single station or substation of an array, referred to by its element name (such as YKR8), as opposed to its array name (YKA in this example).

energy

Occurrence that displays characteristics indicative of a possible nuclear weapons test.

envelope

Overall waveform immediately following a major phase, usually P or S.

event

Unique source of seismic, hydroacoustic, or infrasonic wave energy that is limited in both time and space.

event definition criteria

Rules that specify the minimum number and type of defining phases or stations that are required to define an event.

Glossary ▼

Executive Summary

Product that provides summary statistics regarding the number of S/H/I events formed by automated and reviewed processing, the numbers of those events in the various event-screening categories, the number of Level 4 and 5 radionuclide events, the status of IMS stations, the status of communications, and IDC systems status.

F**false event**

Term used to describe events that are not real or have been built by associating noise or nonseismic detections.

FFT

Fast Fourier Transform.

FIR

Finite Impulse Response.

first motion

First discernible displacement of a seismogram caused by a seismic arrival. Its polarity may or may not be visible.

f-k

Frequency versus wavenumber (k) analysis that maps phase power from an array as a function of azimuth and slowness.

f-k beam

Coherent beam steered to the azimuth and slowness of the tallest peak in a plot of f-k power.

FRF

Finite Response Filter.

G**GA**

Global Association application. GA associates S/H/I phases to events.

GSETT-2

Group of Scientific Experts Second Technical Test.

GSETT-3

Group of Scientific Experts Third Technical Test.

H**hydroacoustic**

Pertaining to sound in the ocean.

Hz

Hertz.

I**IDC**

International Data Centre.

IIR

Infinite Impulse Response (filters also referred to as recursive filters).

IMS

International Monitoring System.

infrasonic

Pertaining to low-frequency (sub-audible) sound in the atmosphere.

K**km**

Kilometer.

L**local**

(1) (distance) Source to seismometer separations of a few degrees or less. (2) (event) Recorded at distances where the first P and S waves from shallow events have traveled along direct paths within the crust.

LP

Long period.

M**m**

1) Meter. 2) Megabyte. 1,024 kilobytes.

m_b

Magnitude of a seismic body wave.

mbmle

Magnitude of an event based on maximum likelihood estimation using seismic body waves.

M_L

Magnitude based on waves measured near the source.

MLE

Maximum Likelihood Estimate.

monitoring system

See IMS and RMS.

M_s

Magnitude of seismic surface waves.

msmle

Magnitude of an event based on maximum likelihood estimation using surface waves.

N**NDC**

National Data Center.

noise

Incoherent natural or artificial perturbations of the waveform trace caused by ice, animals migrations, cultural activity, equipment malfunctions or interruption of satellite communication, or ambient background movements.

nondefining

Arrival attribute, such as arrival time, azimuth, or slowness, which is associated, but not used in calculating the event's location or magnitude.

nondefining phase

Associated phase for which features are not used in the estimating the location and origin time of an S/H/I event.

NSE

Noise Spectrum Equalization.

Glossary ▼

O

octave

Factor of two in frequency (Hz).

1-C

One-component; a single component of motion.

onset

First appearance of a seismic or acoustic signal on a waveform.

origin

Hypothesized time and location of a seismic, hydroacoustic, or infrasonic event. Any event may have many origins. Characteristics such as magnitudes and error estimates may be associated with an origin.

P

parameter

Quantitative attribute of a seismic arrival, such as azimuth, slowness, period, and amplitude.

parameter (par) file

ASCII file containing values for parameters of a program. Par files are used to replace command line arguments. The file is formatted as a list of [token = value] strings.

per

Signal period.

period

Average duration of one cycle of a phase, in seconds per cycle.

phase

Arrival that is identified based on its path through the earth.

phase name

Name assigned to a phase.

PIDC

Prototype International Data Centre.

pipeline

The flow of data at the IDC, from the receipt of communications to the final automated processed data, before the review by the analysts.

polarity

Direction of first motion on a seismogram; either up (compression) or down (dilatation or relaxation).

polarity reversal

Occurrence of depth-phase waveforms that are mirror images of the initial P-type phases

polarization

Form of three-component analysis used to derive azimuth and slowness information from non-array stations.

Preparatory Commission

Preparatory Commission for the CTBTO; new international body funded by States Parties to prepare for implementation of the Treaty. This body will become the CTBTO after entry-into-force of the Treaty.

primary phase

First arriving phase recorded at a S/H/I station.

primary seismic

IMS seismic station(s) or data that is (are) part of the detection network.

Q**quefrency**

Time-delay axis with units of seconds for a cepstrum.

R**REB**

Reviewed Event Bulletin; the bulletin formed of all events that have passed analyst inspection and quality assurance review. The REB runs 48 hours behind real time. The CTBTO is changing the name of this list to SEL3.

reference channel

Array element to which the station's timing is referenced with respect to its other elements, reflecting the timing of the array as a whole. This channel is typically either the element at the center of a circular array or the element at the intersection of a cross-shaped array.

regional

(1) (distance) Source to seismometer separations between a few degrees and 20 degrees. (2) (event) Recorded at distances where the first P and S waves from shallow events have traveled along paths through the uppermost mantle.

residual

Difference in time, azimuth, or slowness between a calculated attribute and its corresponding theoretical value.

rms

Root mean square.

RMS

Radionuclide Monitoring System; the part of the IMS that monitors the atmosphere for radionuclides.

S**s**

Second (time).

SAIC

Science Applications International Corporation.

sample

Any physical entity counted on a detector.

SASC

Slowness-Azimuth Station Corrections.

save

Store an analyzed event to the final database, thereby preventing further changes to the event.

scan

Systematically view all waveforms and seek out possible events missed by the automated system.

Glossary ▼

schema

Database structure description.

Scheme

The language by which ARS and other tools are configured.

s.d.

Standard deviations.

SEB

Standard Event Bulletin; a list of analyst reviewed S/H/I events and event parameters (origin and associated arrival information). The SEB is similar to the REB, but also includes event characterization parameters and event screening results for each event.

secondary phases

Phases that arrive after the primary phase.

secondary phases

Phases that arrive after the primary phase.

select

To choose an element on the screen by clicking on it with the mouse pointer.

SEL1

Standard Event List 1; the bulletin created by total automatic analysis of continuous timeseries data. Typically, the list runs one hour behind real time.

SEL2

Standard Event List 2; the bulletin created by totally automatic analysis of both continuous data and segments of data specifically down-loaded from stations of the auxiliary seismic network. Typically, the list runs five hours behind real time.

SEL3

Standard Event List 3; S/H/I bulletin created by totally automatic analysis of both continuous data and segments of data specifically down-loaded from stations of the auxiliary seismic network. Typically, the list runs 12 hours behind real time.

S/H/I

Seismic, hydroacoustic, and infrasonic.

signature

Appearance of a seismic signal that is unique to the source.

slowness

Inverse of velocity, in seconds/degree; a large slowness has a low velocity.

SLSD

Standard List of Signal Detections.

snr

Signal-to-noise ratio.

SOFAR

Sound Fixing and Ranging. An oceanic waveguide controlled by temperature and salinity.

SP

Short period.

spectrum

A plot of the energy contained in waveforms as a function of frequency.

SRST

Source Region Station Time correction (can be used in S/H/I event location).

SSEB

Standard Screened Event Bulletin; similar in content and format to the Standard Event Bulletin (SEB), but does not include events that were screened out by a standard set of event screening criteria.

SSSC

Source Specific Station Correction (can be used in S/H/I event location).

STA/LTA

Short-term average/long-term average ratio.

StaPro

Station Processing application for S/H/I data.

States Parties

Treaty user group who will operate their own or cooperative facilities, which may be NDCs.

station

Site where a monitoring instrument is installed. Stations can either be single sites (for example, BGCA) or arrays (for example, ASAR).

stepout

Time between two phases, such as pP and P, at a specific station's distance. If the stepout increases as the distance increases, it can be used to identify the phase.

SVD

Singular Value Decomposition.

T**teleseismic**

(1) (distance) Source to seismometer separations of 20 degrees or more. (2) (event) Recorded at distances where the first P and S waves from shallow events have traveled paths through the mantle/core.

theoretical

Point where an arrival is expected to appear on a waveform, based on an event's location and depth.

3-C

Three component.

timing error

Deviation from absolute time, as measured from a station.

Treaty

Comprehensive Nuclear Test-Ban Treaty (CTBT).

Glossary ▼**U****UNIX**

Trade name of the operating system used by the Sun workstations.

UTC

Universal Coordinated Time.

W**waveform**

Time-domain signal data from a sensor (the voltage output) where the voltage has been converted to a digital count (which is monotonic with the amplitude of the stimulus to which the sensor responds).

wavenumber

Vector, k , in the direction of a propagating wave whose magnitude is given by the inverse wavelength of the wave scaled by a factor of 2π .

Index

A

- aftershock scanning [151](#)
- AIC (see Akaike information criterion)
- Akaike information criterion [22](#)
 - algorithm [203](#)
 - method [24](#)
- algorithms [175](#)
- alternate regional group hypotheses [91](#)
- amplitude
 - of regional phases [165](#)
- amplitude and period
 - optimizing seismic parameters [29](#)
 - seismic [27](#)
- analysis
 - enhancing signals [155](#)
 - techniques [155](#)
- analysis (statistics of) [153](#)
- analysis techniques [148](#)
- analyst review overview [146](#)
- analyst reviewed arrivals (tagging for network processing) [110](#)
- anomalous events [151](#)
- array
 - resolution [12](#)
 - response [12](#)
 - seismic [11](#)
- arrival quality test (event definition criteria) [89](#)
- arrival time
 - infrasonic [68](#)

- optimizing seismic parameters [26](#)
 - seismic [22](#)
- ASPM [135](#)
- association
 - corroborating arrivals [84](#)
 - grid-based [82](#)
 - hydroacoustic [57](#)
 - infrasonic [69](#)
 - seismic [33](#)
- association rate [20](#)
- association-based conflict resolution [97](#)
- atmospheric events (analyst review of) [150](#)
- auxiliary arrivals (tagging for network processing) [110](#)
- auxiliary seismic data request parameter optimization [116](#)
- auxiliary seismic station data [113](#)
- azimuth and slowness
 - infrasonic [68](#)
 - optimizing seismic parameters [32](#)
 - prediction [127](#)
 - seismic [30](#)
 - use in location [124](#)

B

- Bayesian analysis [40](#)
 - algorithm [232](#)
- beam [13](#)
 - generation [191](#)
 - incoherent [15](#)
 - optimizing parameters [16](#)
 - response [16](#)
 - steering [14](#)
- beam set [13](#)
 - seismic detection example [17](#)
- best guess location [125](#)

Index ▼

blockage of hydroacoustic phases [136](#)
bubble pulse [55](#)

C

cepstral analysis [218](#)
cepstral features (hydroacoustic) [54](#)
channel jumping [21](#)
chi-square test for arrival compatibility [85](#)
 algorithm [247](#)
cluster analysis (applied to conflict resolution) [96](#)
compatibility test [38](#)
complexity of teleseismic P-waves [165](#)
confidence ellipses [143](#)
conflict resolution
 in network processing [96](#)
 procedure [101](#)
coverage ellipses [143](#)
Cramer-Rao variance estimation [212](#)
cross-correlation functions [193](#)

D

defining phases [123](#)
depth grid cells [80](#)
detecting channel [19](#)
detecting state [19](#)
detection
 hydroacoustic [46](#)
 infrasonic [64](#)
 of hydroacoustic signals [48](#)
 of infrasonic signals [66](#)
 of seismic signals [11](#), [19](#)
 optimizing infrasonic parameters [67](#)
 optimizing seismic parameters for [20](#)
driver arrival [76](#)
 identifying [82](#)
 slowness constraints [83](#)

E

energy ratio [167](#)
error ellipses [143](#)
event categorization [157](#)
event characteristics [162](#)
 first motion [169](#)
 m_b [162](#)
 M_L [163](#)
 M_s [164](#)
 regional phase amplitudes [165](#)
 short-period/long-period energy ratio [167](#)
 spectral and cepstral [167](#)
 teleseismic P-wave complexity [165](#)
 third moment of frequency [168](#)
 time-frequency [169](#)
event confirmation criteria [94](#)
event consistency check [107](#)
event definition criteria [86](#)
 for REB, SEB, and SSEB [153](#)
event location overview [120](#)
event quality (applied to conflict resolution) [98](#)
event screening
 combined score [246](#)
 depth algorithm [243](#)
 $M_s:m_b$ [245](#)
 statistics [243](#)
Executive Summary [2](#), [6](#), [160](#)

F

features [22](#)
filtering
 algorithms [177](#)
 FIR algorithm [185](#)
 IIR algorithm [177](#)
 in analysis [155](#)
 of hydroacoustic data [47](#)
 of infrasonic data [65](#)

- of seismic 3-C data [18](#)
 - of seismic array data [15](#)
- finite impulse response filters [185](#)
- first motion [169](#)
- frequency filters [177](#)
- frequency-wavenumber analysis
 - algorithm [206](#)
 - error estimation [209](#)
 - for infrasonic azimuth and slowness [68](#)
 - for seismic azimuth and slowness [30](#)
 - in interactive analysis [155](#)
- F-statistic
 - algorithm [196](#)
 - for infrasonic signal detection [66](#)
 - for seismic azimuth and slowness error [31](#)

G

- generator [40](#)
- goodness-of-fit (applied to conflict resolution) [98](#)
- grid cells
 - and hydroacoustic phases [81](#)
 - and infrasonic phases [82](#)
 - contents [80](#)
 - depth [80](#)
 - surface [78](#)
- grid development [78](#)

H

- hydroacoustic arrival quality test (event definition criteria) [88](#)
- hydroacoustic arrivals (tagging for network processing) [110](#)
- hydroacoustic blockage [95](#)
- hydroacoustic phases [46](#)
- hydroacoustic station processing steps [45](#)

I

- IASPEI91 [128](#)
- incoherent beam [15](#)
- infinite impulse response filters [177](#)
- infrasonic arrival quality test (event definition criteria) [88](#)
- infrasonic station processing steps [62](#)
- intensity average time (hydroacoustic) [53](#)
- interactive analysis process [147](#)
- interpolation [186](#)
 - closest sample algorithm [186](#)
 - cubic algorithm [187](#)
 - linear algorithm [187](#)
 - two-dimensional algorithm [188](#)

K

- kurtosis (hydroacoustic) [54](#)

L

- large events [89](#)
- late arriving data (tagging for network processing) [109](#)
- location
 - applications that use library [122](#)
 - convergence tests [141](#)
 - divergence tests [141](#)
 - estimating coverage ellipses [143](#)
 - in seismic station processing [41](#)
 - inverting data [138](#)
 - maximum iteration test [142](#)
 - selection of damping [139](#)
 - stability [140](#)
- location analysis [93](#)

M

- magnitudes [162](#)

Index ▼

- algorithms [233](#)
- m_b [234](#)
- M_L [237](#)
- M_s [239](#)
- maximum coherence trace (infrasonic) [69](#)
- maximum likelihood magnitude
 - m_b [235](#)
 - M_s [240](#)
- m_b [162](#)
 - algorithm [234](#)
- missed events (scanning for) [152](#)
- missed rate [20](#)
- M_L [163](#)
 - algorithm [237](#)
- mode propagation calculations [135](#)
- modified Cramer-Rao variance estimation [212](#)
- M_s [164](#)
 - algorithm [239](#)

N

- network average magnitude
 - m_b [235](#)
 - M_L [238](#)
 - M_s [239](#)
- network processing
 - overview [74](#)
 - pipelines [77](#)
- neural network
 - algorithm [229](#)
 - hydroacoustic [57](#)
 - seismic [33](#)
 - training hydroacoustic [59](#)
 - training seismic [36](#)
- noise
 - cultural [156](#)
 - from instrumentation [157](#)
 - seismic [33](#)
- noise spectrum equalization [225](#)

- normalized cross-correlation functions [193](#)
- number of threshold crossings (hydroacoustic) [57](#)

O

- oceanic events (analyst review of) [149](#)
- onset time (hydroacoustic) [50](#)
- outlier analysis [93](#)

P

- peak energy (hydroacoustic) [53](#)
- peak time (hydroacoustic) [53](#)
- phase association (see association)
- phase identification
 - hydroacoustic [60](#)
 - in interactive analysis [156](#)
 - infrasonic [71](#)
 - seismic [37](#)
 - testing for compatibility [38](#)
- phase prediction [106](#)
 - hydroacoustic phase blockage [107](#)
- phase selection (for location) [123](#)
- phase weights for event definition
 - REB, SEB, and SSEB [154](#)
 - SEL1, SEL2 and SEL3 [86](#)
- polarization analysis
 - algorithm [215](#)
 - for seismic 3-C azimuth and slowness [32](#)
- post-analysis processing overview [160](#)
- probability of detection
 - algorithm [240](#)
 - test [94](#)
- probability weighted time
 - algorithm [227](#)
- probability-weighted time (hydroacoustic) [52](#)
- processing sequence [4](#)

Q**QC**

- of analyst reviewed events [151](#), [152](#)
- of hydroacoustic data [46](#)
- of infrasonic data [63](#)
- of REB [152](#)
- of seismic data [10](#)

R

- ranking auxiliary seismic stations [114](#)
- REB (see Reviewed Event Bulletin)
- redundancy
 - in network processing [90](#)
 - testing [94](#)
- redundant event hypotheses [91](#)
- regional phase amplitudes [165](#)
- requesting auxiliary seismic data [115](#)
- Reviewed Event Bulletin [2](#), [6](#)

S

- SASC [32](#)
- scanning for aftershocks [151](#)
- scanning for missed events [152](#)
- screening
 - by depth [170](#)
 - by $M_s:m_b$ [171](#)
 - category definitions [172](#)
 - combining scores [172](#)
- SEB (see Standard Event Bulletin)
- seed event [76](#)
- seed locations [125](#)
- seismic arrival quality test (event definition criteria) [86](#)
- SEL (see Standard Event Lists)
- selecting an event for analyst review [148](#)
- short-period/long-period energy ratio [167](#)
- signal characteristics

- hydroacoustic [49](#)
- infrasonic [68](#)
- seismic [22](#)
- signal grouping
 - hydroacoustic [60](#)
 - infrasonic [71](#)
 - local seismic events [39](#)
 - regional seismic events [39](#)
 - seismic [37](#)
 - teleseismic events [38](#)
 - testing for compatibility [38](#)
- signal summation time (hydroacoustic) [52](#)
- signal time spread (hydroacoustic) [54](#)
- signal type
 - hydroacoustic [57](#)
 - infrasonic [69](#)
 - optimizing hydroacoustic parameters [59](#)
 - optimizing hydroacoustic rules [60](#)
 - optimizing seismic parameters [35](#)
 - optimizing seismic rules [37](#)
 - rules for determining hydroacoustic [58](#)
 - rules for determining seismic [34](#)
 - seismic [33](#)
- skewness (hydroacoustic) [54](#)
- slowness
 - prediction for seismic phases [132](#)
- slowness constraints [83](#)
- slowness planes [64](#)
- slowness/azimuth station corrections
 - for infrasonic phases [138](#)
 - for seismic phases [132](#)
- SLSD (see Standard List of Signal Detections)
- spatial coherence of infrasonic data [64](#)
- spectral analysis
 - noise equalization [225](#)
- spectral and cepstral characteristics [167](#)
- splits in network processing [90](#)
- SSEB (see Standard Screened Event Bulletin)
- STA/LTA
 - algorithms [198](#)
 - for hydroacoustic signal detection [48](#), [198](#)

Index ▼

- for infrasonic signal detection [66](#), [201](#)
 - for seismic signal detection [19](#), [198](#)
 - threshold [20](#)
- Standard Event Bulletin [2](#), [6](#), [160](#)
- Standard Event Lists [2](#), [74](#)
- standard event screening criteria [170](#)
- Standard List of Signal Detections [4](#), [8](#), [44](#), [62](#)
- Standard Screened Event Bulletin [2](#), [6](#), [160](#)
- station association groups [37](#)
- station magnitude
 - m_b [234](#)
 - M_L [237](#)
 - M_s [239](#)
- station processing
 - overview of hydroacoustic [44](#)
 - overview of infrasonic [62](#)
 - overview of seismic [8](#)
- station-specific slowness-azimuth
 - corrections [32](#)
- Stratospheric channel [70](#)
- surface grid cells [78](#)
- system functional model [3](#)

T

- tagging arrivals [109](#)
- teleseismic P-wave complexity [165](#)
- termination time (hydroacoustic) [50](#)
- terrestrial events (analyst review of) [149](#)
- Thermospheric channel [70](#)
- third moment of frequency [168](#)
- three-component
 - detection channels [18](#)
 - snr improvement [18](#)
- time windowing and time stepping of
 - network processing pipelines [111](#)
- time-frequency characteristics [169](#)
- total energy (hydroacoustic) [52](#)
- travel time
 - tables for infrasonic phases [138](#)

- travel-time
 - bulk static station correction [130](#)
 - corrections to seismic phases [128](#)
 - elevation correction [130](#)
 - ellipticity correction [129](#)
 - prediction for seismic phases [127](#)
 - predictions for long period seismic
 - phases LR and LQ [129](#)
 - source-specific station corrections [130](#)
 - tables for hydroacoustic phases [135](#)
 - tables for seismic phases [128](#)
- Tropospheric channel [70](#)

W

- weighted-count test (event definition
 - criteria) [86](#)